DOI:
10.1016/j.envres.2017.10.002

Document Version
Peer reviewed version

Land use regression models for the oxidative potential of fine particles ($PM_{2.5}$) in five European areas.

John Gulliver[1*], David Morley[1], Chrissi Dunster[2], Adrienne McCrea[1], Erik van Nunen,[3] Ming-Yi Tsai[4,5], Nicole Probst-Hensch[4,5], Marloes Eeftens[4,5], Medea Imboden[4,5], Regina Ducret-Stich[4,5], Alessio Naccarati[6], Claudia Galassi[7], Andrea Ranzi[8], Mark Nieuwenhuijsen[9,10,11], Ariadna Curto[9,10,11], David Donaire-Gonzalez[9,10,11], Marta Cirach[9,10,11], Roel Vermeulen,[3] Paolo Vineis[1], Gerard Hoek[3], Frank J Kelly[2].

1. MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom.
2. MRC-PHE Centre for Environment and Health, Environmental Research Group (ERG), King's College London, London.
3. Institute for Risk Assessment Sciences (IRAS), division of Environmental Epidemiology (EEPI), Utrecht University, Utrecht, the Netherlands.
4. Swiss Tropical and Public Health (TPH) Institute, University of Basel, Basel, Switzerland.
5. University of Basel, Basel, Switzerland.
6. Human Genetics Foundation, Turin, Italy.
7. Unit of Cancer Epidemiology, Citta' della Salute e della Scienza University Hospital and Centre for Cancer Prevention, Turin, Italy.
8. Environmental Health Reference Centre, Regional Agency for Prevention, Environment and Energy of Emilia-Romagna, Modena, Italy.
9. ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain.
10. Department of Experimental and Health Sciences, Pompeu Fabra University (UPF), Barcelona, Spain.
11. CIBER Epidemiologia y Salud Pública (CIBERESP), Barcelona, Spain.

* Corresponding author

John Gulliver

MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London, W2 1PG.

Tel: +442075943329; e-mail address: j.gulliver@imperial.ac.uk

ABSTRACT

Oxidative potential (OP) of particulate matter (PM) is proposed as a biologically-relevant exposure metric for studies of air pollution and health. We aimed to evaluate the spatial variability of the OP of measured $PM_{2.5}$ using ascorbate (AA) and (reduced) glutathione (GSH), and develop land use regression (LUR) models to explain this spatial variability. We estimated annual average values (m^{-3}) of OP^{AA} and OP^{GSH} for five areas (Basel, CH; Catalonia, ES; London-Oxford, UK (no OP^{GSH}); the Netherlands; and Turin, IT) using $PM_{2.5}$ filters. OP^{AA} and OP^{GSH} LUR models were developed using all monitoring sites, separately for each area and combined-areas. The same variables were then used in repeated sub-sampling of monitoring sites to test sensitivity of variable selection; new variables were offered where variables were excluded ($p > 0.1$). On average, measurements of OP^{AA} and OP^{GSH} were moderately correlated (maximum Pearson's $R = 0.7$) with $PM_{2.5}$ and other metrics ($PM_{2.5}$absorbance, NO_2, Cu, Fe). HOV (hold-out validation) R^2 for OP^{AA} models was 0.21, 0.58, 0.45, 0.53, and 0.13 for Basel, Catalonia, London-Oxford, the Netherlands and Turin respectively. For OP^{GSH}, the only model achieving at least moderate performance was for the Netherlands ($R^2 = 0.31$). Combined models for OP^{AA} and OP^{GSH} were largely explained by study area with weak local predictors of intra-area contrasts; we therefore do not endorse them for use in epidemiologic studies. Given the moderate correlation of OP^{AA} with other pollutants, the three reasonably performing LUR models for OP^{AA} could be used independently of other pollutant metrics in epidemiological studies.

KEYWORDS

oxidative potential, land use regression, LUR, spatial variability, exposure assessment, air pollution.

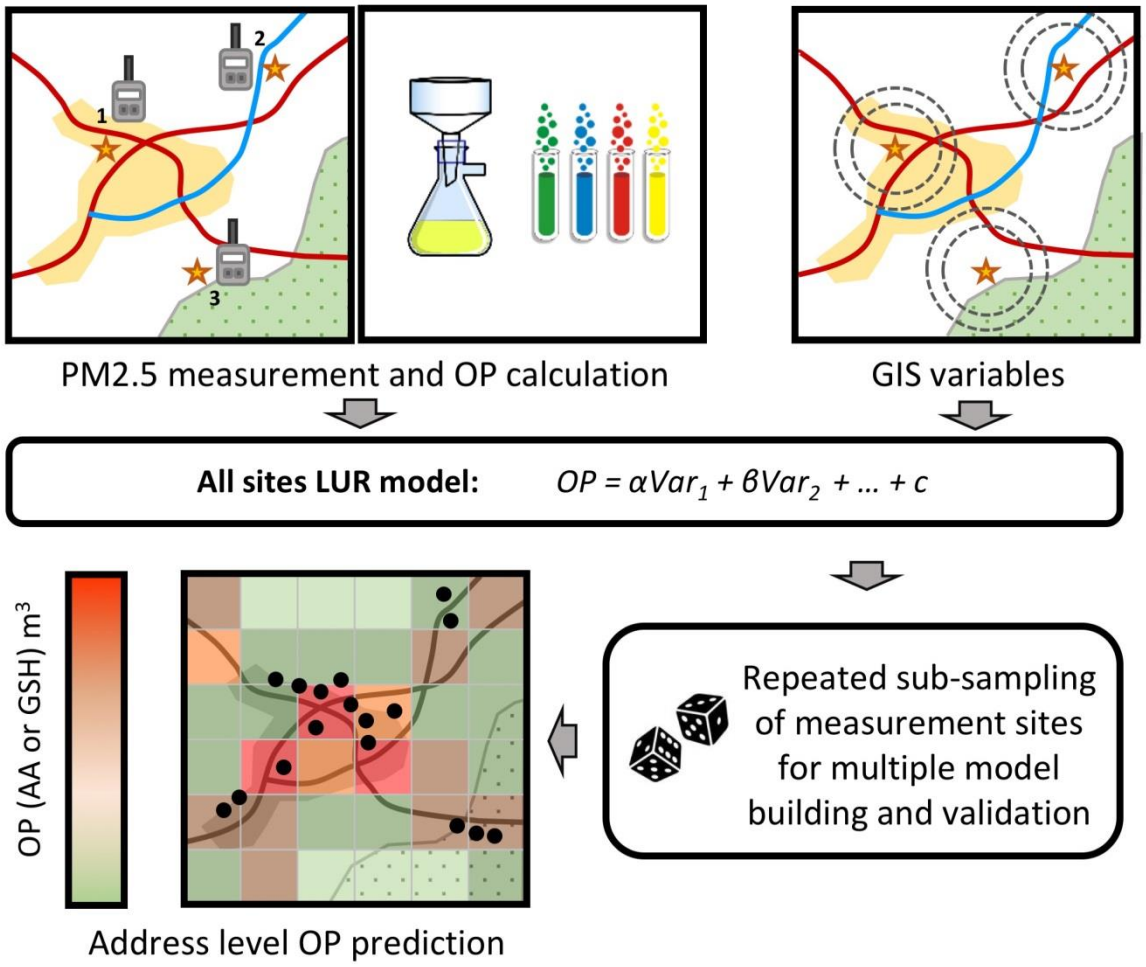
FUNDING

The research was supported by funding from the European Community's Seventh Framework Program EXPOsOMICS study under grant agreement number FP7 308610.

HIGHLIGHTS

- OP moderately correlated (r ~0.6) with $PM_{2.5}$ mass, NO_2, and elemental constituents
- OP^{AA} and OP^{GSH} LUR models were developed and evaluated for five areas
- Three reasonably performing (R^2: 0.45-0.58) OP^{AA} LUR models in hold-out validation
- OP^{GSH} models were not robust to new coefficients derived against sub-samples of sites

GRAPHICAL ABSTRACT



INTRODUCTION

Ambient air pollution is a mixture of gases, organic and non-organic particles, and liquid droplets small enough to remain airborne. Particulates <2.5µm ($PM_{2.5}$) and <10µm (PM_{10}) in diameter has widely been associated with a range of health effects (Brunekreef et al., 2002; Pope et al., 2006; Royal College of Physicians and Royal College of Paediatricians and Child Health, 2016). $PM_{2.5}$ is small enough when inhaled to enter the deeper regions of lung, and has the potential to oxidize the antioxidants that reside in the respiratory tract lining fluid (RTLF) on the surface of the lung (Borm et al., 2007). *In-vitro* models investigating the oxidative potential (OP) of PM have been established in recent years to observe consumption of antioxidants, oxidization of biomolecules (e.g. proteins, DNA, fatty acids) and as a consequence, the capacity to elicit health effects (Ayres et al., 2008). Environmental models of the consumption of antioxidants are related to respirable $PM_{2.5}$ that had been collected from air pollution monitors at different sites types (e.g. major road (street), industrial, urban background, rural) (Boogaard et al., 2012; Janssen et al., 2014; Kunzli et al., 2006; Shi et al., 2006). By combining a synthetic representation of antioxidants in the RTLF with diluted $PM_{2.5}$ in suspension it is possible to observe the depletion of antioxidants across different monitoring sites and relate the OP of $PM_{2.5}$ to different sources including road traffic (Bates et al., 2015; Janssen et al., 2014; Yang et al., 2015a; Yanosky et al., 2012) and biomass burning (Bates et al., 2015).

A range of assays have been used to study the OP or oxidative burden of $PM_{2.5}$ including ascorbic acid (AA) (Fang et al., 2016; Janssen et al., 2014; Maikawa et al., 2016; Weichenthal et al., 2016), antioxidant-reduced glutathione (GSH) (Maikawa et al., 2016; Weichenthal et al., 2016; Yanosky et al., 2012), the consumption of dithiothreitol (DTT) (Bates et al., 2015; Janssen et al., 2014; Jedynska et al., 2017; Yang et al., 2015a), and electron spin resonance (ESR) (Janssen et al., 2014; Yang et al., 2015a). A small number of epidemiologic analyses have shown oxidative burden of $PM_{2.5}$ to be more strongly related than $PM_{2.5}$ mass concentration to a range of outcomes, in both time-series and long-term studies. In Atlanta, USA, (Bates et al., 2015) values of OP^{DTT} were estimated from a time-series of $PM_{2.5}$ samples from a single monitoring site. Regression models created to explain variability in OP^{DTT} included predictor variables for light duty gasoline vehicles, heavy-duty diesel vehicles, and biomass burning. In an epidemiologic analysis of emergency hospital admissions (Fang et al., 2016), OP^{DTT} (in a two pollutant model with OP^{DTT} and total $PM_{2.5}$) was significantly associated with asthma/wheeze (1.015 [CI: 1.002-1.027] per IQR increase) and heart failure (1.024 [CI: 1.004-1.044] per IQR increase); no significant associates were found for $PM_{2.5}$. In Montreal, Canada, the OP of $PM_{2.5}$ (AA and GSH) was determined from

$PM_{2.5}$ personal exposure samples of 62 asthmatic school-aged children collected over 10 days (Maikawa et al, 2016). OP^{GSH} exposure in the previous 24 hours was positively associated (6% increase per IQR change in OP^{GSH}) with fractional exhaled nitric oxide as an indicator of airway inflammation. Weichenthal et al., (2016) derived long-term values of OP^{AA} and OP^{GSH} (% depletion / μg) of $PM_{2.5}$ for 30 provincial monitoring sites across Canada to study oxidative burden of $PM_{2.5}$ and the risk of cause-specific mortality in the CanCHEC cohort. Exposures to the OP of $PM_{2.5}$ were assigned to individuals living within 5km of a monitoring site. For lung cancer, OP^{GSH} was associated with a 12% (95% CI: 5-19) increased risk of mortality compared to a 5% (95% CI: 0.1-10) increased risk for $PM_{2.5}$ mass concentration.

Modelling is commonly used to reveal the spatial contrasts in exposures that cannot be determined from monitoring sites. Unlike other measured pollutant metrics (e.g. NO_2, NO_X, $PM_{2.5}$, O_3, SO_2), it is not possible to deterministically model (e.g. dispersion modelling) OP due to a lack of information on source emissions. Alternatively, land use regression modelling (LUR) - using univariate or multiple regression to establish a relationship of geographical predictors (e.g. road traffic, land use, population distribution) and measured pollutant concentrations (Hoek et al., 2008) - has potential for modelling spatial contrasts in OP. LUR models have been widely used to estimate exposures to regulatory pollutants such as NO_2 and $PM_{2.5}$ (Aguilera et al., 2015; Amini et al., 2014; Beelen et al., 2013; Eeftens et al., 2012a; Henderson et al., 2007; Hoek et al., 2008; Liu et al., 2016) and have been used to produce models for novel metrics such as ultra-fine particles (Abernethy et al., 2013; Hankey et al., 2015; Hoek et al., 2011; Montagne et al., 2015; Rivera et al., 2012; Wolf et al., 2017). LUR models for the OP of PM are less common, but have emerged in Europe in the last few years, and have been developed for different assays including GSH (Yanosky et al., 2012) DTT (Jedynska et al., 2017; Yang et al., 2015), and ESR (Yang et al., 2015).

In a multi-area study on the 'exposome' of air pollution (www.exposomicsproject.eu) (Vineis et al., 2016) we aimed to evaluate the spatial variability of the OP of measured $PM_{2.5}$ (on filters from ESCAPE [Eeftens et al., 2012b] and SAPALIDA [Eeftens et al., 2015] projects) using AA and GSH. Subsequently we aimed to develop OP^{AA} and OP^{GSH} LUR models for each area, and combined-areas models, to explain this spatial variability, and use the LUR models to estimate exposures to OP of $PM_{2.5}$ for cohorts in the EXPOsOMICS project. In order to assess the extent to which OP of $PM_{2.5}$ is an independent metric, we also aimed to assess the correlation of each metric (OP^{AA} and OP^{GSH}) with other pollutants measurements at the same sites, including $PM_{2.5}$ mass concentration, NO_2, and elemental constituents (Cu, Fe, K, Ni, S, Si, V, Zn).

Due to the logistics of establishing a monitoring network with limited monitoring equipment, there often are relatively few sites per study area to develop a LUR model, especially in geographically-wide and multi-center cohort studies (Beelen et al., 2013; de Hoogh et al., 2013; Eeftens et al., 2012a; Tsai et al., 2015). In studies with a relatively low number of monitoring sites there has been a tendency to use all sites to develop a single model for each area and evaluate model performance with leave-one-out-cross-validation (LOOCV) (Beelen et al., 2013; de Hoogh et al., 2013; Eeftens et al., 2012a' Henderson et al., 2007; Zhang et al., 2015). Model performance has been shown to be weaker with low numbers of monitoring sites and the robustness of predictor variables chosen in a single model has been questioned (Basagaña et al., 2012; Wang et al., 2012). To address the issue of low numbers of monitoring sites (e.g. 20) in some of these locations, we also aimed to test the robustness of variable selection and variability in performance of the models. The aim was to repeatedly and randomly select a sub-sample of the monitoring sites to recalculate the coefficients of the models and see if any variables become statistically insignificant.

MATERIALS AND METHODS

$PM_{2.5}$ measurements

$PM_{2.5}$ sampling using Harvard impactors with Teflon Filters (SKC Inc., USA) has previously been described (Eeftens et al, 2012b; Eeftens et al., 2015). In brief, measurements took place for three, two-week periods (summer, winter, intermediate season) over one year: Basel (05/11-03/12), Catalonia (01/09 - 01/10), the Netherlands (whole country) (02/09 - 02/10), London-Oxford (an area including London, The Thames Valley, and Oxford) (01/10 - 01/11), and Turin (02/10 – 01/11). The number of valid filters for each two-week period varied and was dependent on the total of number of sites in each area, typically 4-6 for areas with 20 sites and 8-12 for areas with 40 sites. The mass of $PM_{2.5}$ collected on each filter was subsequently determined and then annual average values of $PM_{2.5}$ mass ($\mu g\ m^{-3}$) were estimated using filters from different seasons; see also Table S1, supporting information. This provided data on $PM_{2.5}$ for between 20 and 40 sites per study area which we were limited to in this study.

Processing of filters

The $PM_{2.5}$ was extracted from the Teflon filters by water-bath sonication into methanol at King's College London. The extracted mass was deduced by weighing of tubes used for the extraction both before and after extraction (Appendix SA.1; supporting information). The Teflon filters from Basel were supplied as half-cuts, and as such were not robust enough for

the sonication extraction procedure. These half-cut filters were placed directly in the synthetic RTLF. The Teflon filters collected in the Netherlands were extracted at the National Institute for public health and the environment (RIVM) (following a similar methanol sonication methodology as that used above) (Yang et al., 2015b).

The $PM_{2.5}$ once extracted from the filter was initially re-suspended to 150 – 500µg/mL and when required for the experimental exposure procedure was diluted to 55.56µg/mL (to provide a final experimental concentration of 50µg/mL). The $PM_{2.5}$ suspensions that displayed very high levels of oxidative activity at 50µg/mL within the RTLF exposure model were further diluted and re-exposed to obtain reliable OP data (< 90% oxidation of antioxidant) at an appropriate concentration of 25, 12.5 or 6.25µg/mL.

Determination of Oxidative Potential

A 50µL aliquot of synthetic RTLF containing equi-molar concentrations of AA and GSH was added to the exposure tubes containing 450µL of the diluted $PM_{2.5}$ in suspension (Appendix SA.2; supporting information). The RTLF, now containing 200µmoles/L of each antioxidant and 50µg/mL $PM_{2.5}$ (or its equivalent 1 in 2 dilution) was incubated for 4 hours at 37^{o}C with constant mixing. In-house controls of particle-free, negative (M120) and positive (urban particulate NIST1648a – NIST, USA) PM, extracted laboratory filter and probe sonication blanks, were incubated in parallel to the $PM_{2.5}$ samples to control for background antioxidant oxidation, delivery of expected oxidation by the –ve and +ve controls in the RTLF exposure model, and for checks of cross-contamination from the laboratory blanks (Appendix SA.3; supporting information). To eliminate as much background antioxidant oxidation as possible from the model system, HPLC-grade water that had been treated previously with Chelex-100 resin (Sigma, London-Oxford) was used throughout for preparation of stocks and dilutions. The RTLF/$PM_{2.5}$ exposure experiments were undertaken, at pH7.0. Immediately following the 4-hour incubation the micro-tubes were centrifuged at 13,000rpm for 1 hour at 4^{o}C, followed by removal of aliquots into 100mM phosphate buffer pH7.5 (for GSH analysis) and 5% v/v *meta*-phosphoric acid (for AA analysis). All tubes were immediately stored at -70^{o}C. A summary of the measured OP (m^{-3}) totals (i.e. the sum of OPP^{AA} and OP^{GSH}) for each country is provided in Appendix SA.4 (supporting information). OP^{TOTAL} m^{-3} was determined from OP^{TOTAL} μg^{-1} (OP^{AA} μg^{-1} + OP^{GSH} μg^{-1}) multiplied by the ambient $PM_{2.5}$ mass concentration (μg m^{-3}).

The number of filters collected at sites in each study area and the number of available values (m^{-3}) of OP^{AA} and OP^{GSH} is shown in Table S2 (supporting information). Values of OP

m^{-3} for AA and GSH were temporally adjusted, following established procedures (Eeftens et al., 2012b; Eeftens et al., 2015) as the basis for calculating annual average (i.e. the average of up to three, two-week measurements) values of OP^{AA} and OP^{GSH} for each site (Appendix SA.5, supporting information). The requirement for annual average OP was valid filter measurements following temporal adjustment for two or more two-week periods representing different seasons. The main source of missing filters (Table S2) was a lack of reference site data (and a relationship too weak between OP^{AA} or OP^{GSH} and $PM_{2.5}$ to impute reference site data) or negative values following temporal adjustment.

To evaluate whether OP^{AA} and OP^{GSH} are useful as independent air pollution metrics for epidemiological studies we assessed their correlation with existing measurements (with the exception of Basel (SAPALDIA) all other measurements came from the ESCAPE study) of $PM_{2.5}$, $PM_{2.5}$absorbance, NO_2, and eight selected elements (Cu, Fe, K, Ni, S, Si, V, Zn) from XRF analysis.

GIS predictor variables

Using a GIS (Geographical Information System) and data from the same years as measurements, predictor variables (Table S3, supporting Information) were generated locally for each measurement site and linked to the annual average values of OP. Predictor variables and buffer sizes were similar to those used in the ESCAPE study (Beelen et al., 2013; Eeftens et al., 2012a). Road traffic predictors were generated within buffers of radii 50, 100, 300, 500, 1000 meters, and from measures of inverse distance from the nearest major road, using the best available local data on road geography and traffic flows. Data on population (European Environment Agency) and land cover (COoRdination of Information on the Environment; CORINE) were generated within buffers of radii 100, 300, 500, 1000, and 5000 meters.

Development of LUR models

We implemented a strategy for development and evaluation of LUR models in response to concerns that having a low number of sites raises doubt about the robustness and generalizability of models (Basagaña et al., 2012; Wang et al., 2012). Models for each study area were developed using the following steps:

1) All sites (i.e. one value of OP^{AA} and OP^{GSH} per site) by area were used to select the set of variables that gave the highest overall adjusted-R^2 (explained variability in measured OP^{AA} and OP^{GSH}), with the proviso that each variable added at least 1% to the adjusted-R^2, values of p for each variable on entry were <0.05 and remained <0.1 with the final set of variables, the pre-defined direction of effect (+ or -) remained

unchanged, and values of variance inflation factor (VIF) were <3. This is similar to the set of rules used in the ESCAPE study.

2) A repeated, random sub-sampling (RSS) procedure was used to create variations of the initial model (i.e. the model from step 1) using all sites, where 90% of sites were used each time to recalculate variable coefficients and 10% of sites were reserved each time (and later pooled) for hold-out (i.e. out-of-sample) validation (HOV).

In step 2) above, the remaining 90% of sites were used to recalculate the coefficients of the all sites LUR model. New variables were only allowed at this stage if variables from the models based on all sites were dropped due to no longer being significant ($p > 0.1$). Monitoring sites were randomly left out (and then replaced for the next iteration of RSS) up to N number of times, which varied depending on the number of monitoring sites (10% of the total number of sites in each case). It was ensured in advance that a site could only be left out a maximum of N times (e.g. 2 times for a sample size of 20; 4 times for a sample size of 40; etc.) over all iteration cycles to reduce possible bias. The iteration process stopped when all sites had been left out N times. This means, for example, that a model developed on 20 sites will result in 40 sites for HOV (i.e. 10% of the total number of sites ($n = 2$) in each of 20 iterations of RSS); values for 10% site selection were rounded to the nearest integer (e.g. 3.9 to 4 in the case of the Netherlands). Values of min, 10^{th}%ile, median, 50^{th}%ile, and maximum R^2 (coefficient of determination), RMSE (root mean squared error), NRMSE (RMSE normalized by mean of measurements) for each model, and p-values for each variable from each model, were pooled (i.e. to test robustness) across all models. We also evaluated values of Cook's D for each model iteration to identify influential observations ($D>1$) in relation to specific monitoring sites. Finally, model residuals were checked for normality.

Combined area OP models

To develop "all areas" OP models, we combined data on OP^{AA} from all sites and OP^{GSH} from all sites except London-Oxford where no average OP^{GSH} was available. In addition to the procedures for local models we also stratified by study area. We used multiple regression with and without fixed-factors for study area, and subsequently linear mixed-effects regression modelling, specifying random intercepts to account for differences in background concentrations between countries (study areas). We also undertook leave-one-area-out analysis (i.e. iteratively dropping one area from the "all areas" models).

Model evaluation

The 10% of sites left out of each RSS iteration (i.e. HOV) were combined and used for a single, overall evaluation in terms of R^2 (coefficient of determination; i.e. 1-(MSE / variance of observations) and RMSE. Thus, each observation (measurement of OP) was compared with 10% of model predictions for the same site. We also compared HOV following RSS with LOOCV (R^2 and RMSE) on the all sites models as it is commonly used in other studies.

RESULTS

Quality control

Teflon lab blanks and field blanks were included in all areas except Basel (not available) and treated in the same way in the $PM_{2.5}$ suspension as described above for all other filters (Table S1; Appendix SA.1; supporting information). For OP^{GSH} the %CV (coefficient of variation) of analysis was less than 10% with a minimum detection limit of 0.3µmoles/L. For OP^{AA} the %CV of analysis was less than 5% with a minimum detection limit for ascorbic acid of 0.5µmol/L. Values of OP in µmoles/L were converted to µg and subsequently converted to OP concentrations (m^{-3}).

Differences related to study area and type of monitoring site

Median OP^{AA} was more than two-fold higher in Turin (93.1 m^{-3}; SD = 34 m^{-3}) than in Catalonia (44.4 m^{-3}; SD = 31.2 m^{-3}) and the Netherlands (41.0 m^{-3}; SD = 14.1 m^{-3}), and approximately three-fold higher in Turin than in London-Oxford (33.2 m^{-3}; SD = 16.1 m^{-3}) and Basel (28.4 m^{-3}; SD = 5.1 m^{-3} (Figure 1). Median OP^{GSH} was more than two-fold higher in Turin (10.2 m^{-3}; SD = 7.5 m^{-3}) than in Basel (4.7 m^{-3}; SD = 1.3 m^{-3}), and about three-fold higher in Turin than in the Netherlands (3.9 m^{-3}; SD = 1.4 m^{-3}) and Catalonia (3.4 m^{-3}; SD = 4.8 m^{-3}). Although values of OP^{AA} and OP^{GSH} varied by study area ($p < 0.001$), excluding the Turin sites negates any OP^{GSH} variation by study area ($p = 0.39$). In all locations OP^{TOTAL} was dominated by OP^{AA}; median OP^{AA} is approximately 15, 11, 9 and 6 times higher than median OP^{GSH} in Catalonia, the Netherlands, Turin, and Basel (OP^{GSH} not available for London-Oxford), respectively. The large differences in both OP^{AA} and OP^{GSH} between Turin and other locations were not related to differences in the ratio of either metric to $PM_{2.5}$ (µg m^{-3}), which was similar in magnitude for OP^{AA} in Turin, Catalonia and London-Oxford, and similar for OP^{GSH} in Turin, Basel and Catalonia.

Measured OP^{AA} and OP^{GSH} at street (S) sites were on average 1.5 ($p<0.01$) and 1.4 ($p>0.05$) times higher than at urban background (UB) sites, and 2.0 ($p<0.01$) and 2.2 ($p<0.05$) times higher than at regional background (RB), respectively (Table S6, supporting information). Turin is the only area where there was a non-significant difference ($p>0.05$) between S and UB sites, hence the overlap in IQRs for site type (Figure 1). With the exception of the

Netherlands, all ratios of S/UB for OP^{GSH} for individual areas were non-significant ($p>0.05$). Ratios between site types for OP^{AA} and OP^{GSH} were broadly comparable to those for $PM_{2.5}$absorbance and NO_2, whereas Cu and Fe had substantially higher ratios for both S/UB and UB/RB (Table S6; supporting information).

Correlations between measured pollutant metrics

Correlations between OP^{AA} and OP^{GSH} and $PM_{2.5}$, $PM_{2.5}$ absorbance, NO_2, and elemental constituents were highly variable between areas (Table 1). OP^{AA} was on average (of correlations from each area) moderately correlated (0.48) with OP^{GSH}. The average correlations across all areas were strongest but still moderate ($r \sim 0.6$) for OP^{AA} and $PM_{2.5}$, $PM_{2.5}$ absorbance, NO_2, Cu and Fe, being strongest in London-Oxford, and weakest in Basel, Catalonia and Turin. Correlations of OP^{GSH} and other metrics were generally weak, in the region of ~0.3 for $PM_{2.5}$, $PM_{2.5}$absorbance, NO_2, Cu and Fe, and non-significant by individual area, except in the Netherlands. In pooling data from all areas, OP^{AA} explained about 50% of the variability (R^2) ($r \sim 0.7$) in OP^{GSH}, $PM_{2.5}$, $PM_{2.5}$absorbance, NO_2, Cu and Fe.

LUR models

Distributions of OP^{AA} and OP^{GSH} were near-normal so we did not undertake data transformation (e.g. Ln) prior to model development.

OP^{AA} LUR models

Values of R^2 for the model using all sites (Table 2) were 0.44 (Basel), 0.64 (Catalonia), 0.84 (London-Oxford), 0.60 (the Netherlands), and 0.56 (Turin); see also Table S4, supporting information. All models for OP^{AA} included at least one variable on traffic load and/or road length accompanied in some cases by additional variables on population distribution (the Netherlands), urban green space/natural land (Catalonia and Turin), and residential land (Basel). In Catalonia site-specific, fixed factors were included for the reference sites (i.e. describing background concentrations) relating to the three distinctive areas where monitoring sites were located.

OP^{GSH} LUR models

For OP^{GSH} (Table 3 and Table S5, supporting information) values of initial model R^2 were 0.51 (Catalonia), 0.22 (Turin) and 0.44 (the Netherlands). Models included at least one variable on traffic with the addition of semi-natural land (the Netherlands) or industrial land (Turin). It was not possible to develop a statistically significant model for Basel, and for

London-Oxford no model was possible due to the lack of annual average measurements on OP^{GSH}.

Combined area LUR models

Initial combined areas models using linear mixed effects (i.e. random intercepts on a variable defining country) were created for OP^{AA} (R^2 = 0.65) and OP^{GSH} (R^2 = 0.39). Most of the explained variance was due to study area: adjusted R^2 is 0.21 and 0.07 for the respective models without study area. In both cases the main variables are traffic load on major roads within a 50m circular buffer with OP^{AA} having additional variables on length of all roads and semi-natural land. For combined-areas, regression without fixed factors did not yield statistically significant models. Residuals from area-specific and combined models were normally distributed and all area-specific and combined-areas models produced values of Cook's D < 1 with the exception of one monitoring site in Turin.

Repeated sub-sampling

In RSS, median values of R^2 for OP^{AA} (Table 2) and OP^{GSH} (Table 3) were either the same or very close (<5% change) to those from models using all monitoring sites, but using different combinations of sites there was substantial variation in R^2 especially in locations with lower numbers of monitoring sites (e.g. 20). None of the variables in the initial Catalonia model for OP^{GSH} were statistically significant in any combination of monitoring sites in RSS (Table 3). The proportion of values of Cook's D > 1 were very low (<2%) with the exception of the OP^{AA} model for Turin (~6%). These sites did not significantly affect the magnitude of coefficients for the different variables so they were retained. In RSS, most variables selected for initial models remained significant (Figure S1, supporting information) and where variables were dropped there were no new variables or changes to buffer sizes of existing variables. Boxplots of the variability in p-values for individual variables for models in RSS are shown in Figure S2 (supporting information).

Hold-out validation

Compared to values returned in the models using all sites, there was moderate (i.e. <20%) inflation in HOV RMSE for Basel, Catalonia, the Netherlands, and combined-areas OP^{AA} models (Table 2) and all HOV OP^{GSH} models. HOV RMSE increased by 36% and 75% in Turin and London-Oxford, respectively, hence the associated substantial drop in values of HOV R^2 (Table 3). Values of R^2 were within 6-7% (of 100% possible explained variability) of those from the initial model in Catalonia (0.58) and the Netherlands (0.53) and 5% for the combined model (0.60) for OP^{AA}. Substantial reductions in values of R^2 from initial models were seen in Basel (0.21), Turin (0.13), and London-Oxford (0.45). HOV for OP^{GSH} yielded

reduction of the total possible explained variance (in R^2) by 13% in the Netherlands (0.31) and 12% in the combined model (0.27). The model for Turin for OP^{GSH} performed very poorly (R^2 = -0.03) in HOV. Scatterplots of measured versus modelled OP metrics for HOV are shown in Figure S3 (supporting information).

DISCUSSION

Substantial spatial variation in estimated annual average values of OP^{AA} and OP^{GSH} within- and between-site type (S, RB, UB) and between countries was identified. We developed and evaluated OP^{AA} models for five areas but only produced two single area models for OP^{GSH}, one of which (Turin) performed very poorly in HOV. Combined-areas modelling produced models dominated by area effects with weak local predictors. This is the first time LUR models have been developed for OP^{AA} and the second time for OP^{GSH} (Yanosky et al., 2012).

Comparison between measurements of OP by site type and with other metrics

We found ratios of 1.5 and 2.0 for OP^{AA} and 1.4 (but not statistically significant) and 2.2 for OP^{GSH} between S/UB and S/RB sites, respectively, using a much larger and geographically diverse number of measurements sites than have been previously published for OP. In the Netherlands (Yang et al., 2015b), using the same sites and extracts from the same $PM_{2.5}$ filters as in the present study, ratios for S/UB are 1.2 ($p < 0.05$) and 1.4 ($p < 0.01$) for OP^{DTT} and OP^{ESR}, respectively. Our ratios for S/UB are of similar magnitude for the Netherlands (Table S6, supporting information) for both OP^{AA} and OP^{GSH}. Ratios for S/UB are, in contrast, lower (<1.2) for OP^{DTT} in a ten area study across Europe (Jedynska et al., 2017) with some sites showing higher values of OP^{DTT} at UB sites than S sites, with little variance overall in the difference in OP^{DTT} between UB and RB sites. There is a tendency for measurements of OP^{DTT} to have relatively low contrasts between S and UB sites (Janssen et al., 2014; Jedynska et al., 2017; Yang et al., 2015a; Yang et al., 2015b). Other metrics such as OP^{AA} and OP^{ESR} may therefore have greater potential for differentiating pervasive sources of exposures such as road traffic in near-roadway studies. OP^{DTT} may be useful in explaining spatial contrasts in other sources such as biomass burning (Bates et al., 2015; Fang et al., 2016) and may relate better to background PM mass and organic carbon than road traffic components of PM (Fang et al., 2016; Janssen et al., 2014). We and others (Yang et al., 2015b) found larger spatial gradients for OP than PM mass within urban areas. In applying data in an epidemiologic analysis, Weichenthal et al. (2016) found that spatial gradients in $PM_{2.5}$ oxidative burden (OP^{AA} and OP^{GSH}) were higher than for $PM_{2.5}$ mass concentrations.

We found that on average (i.e. the average of correlations from each area) OP explains < 40% (r ~0.6) of the variability in measurements of other metrics (Table 1). In Yang et al. (2015a), based on the same $PM_{2.5}$ samples as used in the present study (n=40), correlations were notably higher (R: 0.72 – 0.92) between OP^{ESR} and some metrics ($PM_{2.5}$, $PM_{2.5}$absorbance, NO_2, Cu, Fe) but not for OP^{DTT} (R < 0.7), and this pattern did not especially change when comparing these metrics in terms of predicted residential exposures from LUR models. In our study, (Table 1), the highest correlations (0.57 - 075) were also between OP^{AA} and $PM_{2.5}$, $PM_{2.5}$absorbance, NO_2, Cu, Fe (Table 1) but they were notably lower than for OP^{ESR}. These results suggest that OP^{AA} and OP^{ESR} in the Netherlands were reacting to different components of $PM_{2.5}$ that result in OP^{AA} being more independent of other pollutants (including $PM_{2.5}$ mass) than OP^{ESR}. Our results also suggest that OP^{AA} should be able to differentiate exposures to traffic-related air pollution and other sources.

LUR model performance

Due to the low number of sites (i.e. 20) in some areas, we repeatedly built different versions of the all sites models using a sub-sample (RSS) of all the measurement sites (N-10%), to test the robustness of variables selected for initial models where we used all sites. K-fold model development and evaluation (i.e. the measurement data are systematically separated into groups of sites, separate models are built for each group, and each model produced is used to predict on the held-out data each time) are not new to LUR modelling (Amini et al., 2014; Gulliver et al., 2013; Wang et al., 2016), but this is the first time such an approach has been used for OP. We used RSS to select groups (not simply splitting the data once into groups) of monitoring sites to increase the number of iterations of models. We chose to use k=10% for N-k in RSS to provide a number of models equivalent to the number of measurement sites. We could have chosen other values of k but felt that a higher proportion of held-out sites would too greatly reduce the N in those areas where there was a low number of total measurements sites (e.g. 20). In RSS, values of R^2 for OP^{AA} varied more in terms of inter-decile range (i.e. absolute difference between the 90^{th}%ile and 10^{th}%ile of values) for areas with 20 sites (Basel = 29%; Turin = 18%) than those with larger numbers of sites (Catalonia = 7%, the Netherlands = 9%), with the exception of the London-Oxford (8%). London-Oxford had, however, the largest range of values of R^2 (39%) that relate to the inclusion/exclusion of one S monitoring site with a substantially higher level of OP^{AA} than other sites (Figure S3, supporting information). For OP^{GSH} model, R^2 was weaker and more variable in terms of the inter-decile range with lower numbers of sites in Turin (16%) than the Netherlands (11%).

The performance of models is thus sensitive to the number of measurement sites and inclusion/exclusion of specific sites consistent with findings of other studies (Basagaña et al., 2012; Wang et al., 2012). We suggest that where monitoring site numbers are low (e.g. 20), RSS could be used to test variable robustness, and information on the variability in model performance (R^2, RMSE) from RSS and HOV can be used to inform the "quality" of exposure in epidemiological studies. Based on RSS and HOV, our models of OP^{AA} worked well in some areas (Catalonia, the Netherlands, London-Oxford) but not in others (Basel and Turin). We recommend using the initial model with all sites and then an average of the permutations (RSS) of the all sites model could be used in epidemiological studies in sensitivity analysis.

We had less success in developing models of OP^{GSH}, being unable to produce statistically significant models for Basel (given the relatively small number of sites and the limited spatial contrasts in OP^{GSH}) and London-Oxford (due to the lack of measurements). Although we produced all-sites models for Catalonia and a combined-areas model, these models became non-significant in RSS and HOV. There is a OP^{GSH} model for the Netherlands but the performance in HOV was moderate ($R^2 = 0.31$). Measurements of OP^{GSH} significantly ($p < 0.05$) differentiated UB from RB sites but not S from UB sites, which may explain, given the localized nature of variables offered in our models, the weaker performance of OP^{GSH} LUR than OP^{AA} LUR. The only other study (Yanosky et al., 2012) to develop a model of OP^{GSH} was in London, UK, based on the amount of GSH lost in a 50 μg mL^{-1} concentration of suspended PM, using PM_{10} filters from TEOM monitors collected in the period 2002 to 2006. The spatial model (annual average based on measurements sites with at least 40 weeks of data) resulted in a cross-validation R^2 of 0.73. The high value of R^2 may relate to the combined benefits of continuous monitoring for 40 weeks, the number of monitoring sites (n=34), large spatial contrasts in OP^{GSH} relative to source activity, the selected predictors they were able to offer (e.g. differentiating between emissions of PM from tailpipe and brake/tire wear), three categories of vehicles (separating light and heavy goods vehicles, and all other vehicles), and errors that are small relative to mean OP^{GSH}.

Our combined-areas models for OP^{AA} and OP^{GSH} had large area effects and a low level of explained variability related to local GIS predictor variables. Even including study area effects, the combined OP^{GSH} model performed relatively poorly in HOV ($r^2 = 0.27$). Based on the differences in intercepts for some countries (Figure S3, supporting information) we also attempted to recreate combined-areas models for both OP^{AA} and OP^{GSH} by iteratively leaving one area out. It was not possible, however, to produce statistically significant models for any combination of areas. A combined OP^{DTT} model produced for ten European areas (Jedynska et al., 2017) using fixed effects on some areas produced $R^2 = 0.26$ in LOOCV.

Performance of combined models could be affected by differences in timing of $PM_{2.5}$ measurements between areas. In our study $PM_{2.5}$ measurements were not all made in the same year: (predominantly in) 2009 for Catalonia and the Netherlands, 2010 for London-Oxford and Turin), 2011 for Basel. Spatial patterns are known to change between years, especially for regional pollutants such as $PM_{2.5}$ (Eeftens et al., 2012b), but we are unable to assess any potential implications of this on our data.

The performance of our models may also be affected by not allowing spatial predictors to initially change in RSS. In RSS we only allowed new variables from the full list (Figure S3) to replace those that were dropped due to being non-significant ($p > 0.1$). Otherwise we did not allow new variables in RSS as this would have caused a further reduction of sites to develop models, meaning only 18 sites for RSS in some areas. This may have resulted in an inability to represent some types of source contributions in OP models (e.g. industrial land which was present only in OP^{GSH} for Turin). Studies (Armiri et al. 2014; van Nunen et al., 2017; Wang et al., 2016) that allowed variables to change in developing multiple models had the advantage of a larger number of sites (> 40). It may also be the case that model performance was compromised by being limited to one reference site for each study area (or each distinctive area in Catalonia). A single reference site may not always be sufficient to provide background values of OP, which may explain why we had negative temporally adjusted values of OP for some filters which resulted in a reduction of sites in some areas (e.g. four sites removed in Catalonia for OP^{GSH}). Values of R^2 and RMSE from LOOCV (Table 2 and Table 3) were notably higher (e.g. 11%-37% for OP^{AA}) than those from HOV (following RSS) in areas with 20 sites but almost the same in areas with >= 39 sites. Furthermore, LOOCV statistics are presented for OP^{GSH} in Catalonia whereas none of the variables remained significant in RSS. This suggests that studies with low numbers of sites may have overestimated model performance if using LOOCV. In reflecting on the performance of models, we reproduced models where the p-value for variable inclusion was relaxed to 0.1. This did not result in improvements, as in a few instances where we were able to produce models with different combinations of variables, LOOCV, RSS and HOV performance was worse than with the original inclusion criteria.

There are a number of other possible reasons for the overall relatively low performance of our models. We noted that OP^{AA} and OP^{GSH} were, at best, moderately correlated to other pollutants for which LUR models have been successfully developed. It may be that OP^{AA} and OP^{GSH} relate to some other sources and atmospheric processes that we have not accounted for in our models such as biomass/wood burning. We were not aware of any significant influence of biomass/wood burning close to sites used in this study, but there may have

been some diffuse emissions from these sources that we were unable to represent in our models. Spatial contrasts of OP^{GSH} are also relatively low, with a higher level of uncertainty (ratio of limit of detection of OP^{GSH} to estimated values of OP^{GSH}) than OP^{AA}, which may have exacerbated the influence of relatively large errors on OP^{GSH} model derivation. We also attempted to create models for OP^{TOTAL} but this did not produce different models than for OP^{AA} as ~90% of measured OP^{TOTAL} is OP^{AA}. The question also arises whether our filter based OP measurements provided a sufficiently precise and reliable measure of the oxidative property of ambient air as do measurements of particle mass or gaseous. If the time between deposition of particles on filters and/or the storage and handling of filters affect the oxidative properties, this may add non-systematic variation to the measures ultimately used in the models. Data quality of GIS variables offered into LUR models is unlikely to explain model performance as they have successfully been used to develop models for other pollutant metrics (Beelen et al. 2013; de Hoogh et al., 2013; Eeftens et al., 2012a).

Comparison with other OP LUR modelling studies

Our work is the first to develop LUR models for OP^{AA}, whilst the only other study (Yanosky et al., 2012) to develop LUR models for OP^{GSH} was a model for 34 sites in London, UK, where traffic variables (NO_X exhaust emissions from heavy goods vehicles within 100m and PM_{10} brake and tire emissions within 50m) were the sole spatial predictors in the model. Information on road traffic also provided the highest partial R^2 of variables included in LUR (40 sites) for OP^{DTT} (0.33 out of a total of 0.55) and OP^{ESR} (0.37 out of a total of 0.64) in the Netherlands (Yang et al., 2015a), with good performance in model evaluation (LOOCV R^2 of 0.47 and 0.60 for OP^{DTT} and OP^{ESR}, respectively). In contrast none of the OP^{DTT} LUR models in five European areas (Athens, Catalonia, the Netherlands, Oslo, Paris) included predictor variables for road traffic (Jedynksa et al., 2017); in an additional five areas it was not possible to develop statistically significant models. Generally poor model performance was attributed to low levels of variability in OP^{DTT}, low numbers of sites in each area (16 in the Netherlands), and a lack of GIS variables specific for OP^{DTT}. This again points to the number of monitoring sites being crucial in model development, hence the need for a methodology, such as RSS applied here, to make an assessment of the robustness of variables included in models where the number of sites is especially low. Nevertheless, in general, LUR performance is not likely to be as good for OP as for pollutants such as NO_2 and $PM_{2.5}$ / PM_{10} where values of LOOCV or HOV R^2 often exceed 0.7 (Beelen et al. 2013; Eeftens et al., 2012a; Liu et al., 2016). OP may, however, have the capability to differentiate exposures for S, UB, and RB sites where valid LUR models can be produced.

REFERENCES

Abernethy, R. C.; Allen, R. W.; McKendry, I. G.; Brauer, M., A land use regression model for ultrafine particles in Vancouver, Canada. *Environ Sci Technol* **2013,** *47* (10), 5217-25.

Aguilera, I.; Eeftens, M.; Meier, R.; Ducret-Stich, R. E.; Schindler, C.; Ineichen, A.; Phuleria, H. C.; Probst-Hensch, N.; Tsai, M.-Y.; Künzli, N., Land use regression models for crustal and traffic-related PM2.5 constituents in four areas of the SAPALDIA study. *Environmental Research* **2015,** *140*, 377-384.

Amini, H.; Taghavi-Shahri, S. M.; Henderson, S. B.; Naddafi, K.; Nabizadeh, R.; Yunesian, M., Land use regression models to estimate the annual and seasonal spatial variability of sulfur dioxide and particulate matter in Tehran, Iran. *Sci Total Environ* **2014,** *488-489*, 343-53.

Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; Künzli, N., Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmospheric Environment* **2012,** *54*, 634-642.

Bates, J. T.; Weber, R. J.; Abrams, J.; Verma, V.; Fang, T.; Klein, M.; Strickland, M. J.; Sarnat, S. E.; Chang, H. H.; Mulholland, J. A.; Tolbert, P. E.; Russell, A. G., Reactive Oxygen Species Generation Linked to Sources of Atmospheric Particulate Matter and Cardiorespiratory Effects. *Environ Sci Technol* **2015,** *49* (22), 13605-12.

Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.-Y.; Künzli, N.; Schikowski, T.; Marcon, A.; Eriksen, K. T.; Raaschou-Nielsen, O.; Stephanou, E.; Patelarou, E.; Lanki, T.; Yli-Tuomi, T.; Declercq, C.; Falq, G.; Stempfelet, M.; Birk, M.; Cyrys, J.; von Klot, S.; Nádor, G.; Varró, M. J.; Dėdelė, A.; Gražulevičienė, R.; Mölter, A.; Lindley, S.; Madsen, C.; Cesaroni, G.; Ranzi, A.; Badaloni, C.; Hoffmann, B.; Nonnemacher, M.; Krämer, U.; Kuhlbusch, T.; Cirach, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Bellander, T.; Korek, M.; Olsson, D.; Strömgren, M.; Dons, E.; Jerrett, M.; Fischer, P.; Wang, M.; Brunekreef, B.; de Hoogh, K., Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmospheric Environment* **2013,** *72*, 10-23.

Boogaard, H.; Janssen, N. A.; Fischer, P. H.; Kos, G. P.; Weijers, E. P.; Cassee, F. R.; van der Zee, S. C.; de Hartog, J. J.; Brunekreef, B.; Hoek, G., Contrasts in oxidative potential and other particulate matter characteristics collected near major streets and background locations. *Environ Health Perspect* **2012,** *120* (2), 185-91.

Borm, P. J.; Kelly, F.; Kunzli, N.; Schins, R. P.; Donaldson, K., Oxidant generation by particulate matter: from biologically effective dose to a promising, novel metric. *Occup Environ Med* **2007,** *64* (2), 73-4.

Brunekreef, B.; Holgate, S. T., Air pollution and health. *Lancet* **2002,** *360* (9341), 1233-42.

de Hoogh, K.; Wang, M.; Adam, M.; Badaloni, C.; Beelen, R.; Birk, M.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; Eeftens, M.; Eriksen, K.; Eriksson, C.; Fischer, P.; Grazuleviciene, R.; Gryparis, A.; Hoffmann, B.; Jerrett, M.; Katsouyanni, K.; Iakovides, M.; Lanki, T.; Lindley, S.; Madsen, C.; Molter, A.; Mosler, G.; Nador, G.; Nieuwenhuijsen, M.; Pershagen, G.; Peters, A.; Phuleria, H.; Probst-Hensch, N.; Raaschou-Nielsen, O.; Quass, U.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Schwarze, P.; Tsai, M. Y.; Yli-Tuomi, T.; Varro, M. J.; Vienneau, D.; Weinmayr, G.; Brunekreef, B.; Hoek, G., Development of land use regression models for particle composition in twenty study areas in Europe. *Environ Sci Technol* **2013,** *47* (11), 5778-86.

Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.; Galassi, C.; Grazuleviciene, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.;

Lanki, T.; Lindley, S.; Madsen, C.; Molter, A.; Nador, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.; Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M. Y.; Yli-Tuomi, T.; Varro, M. J.; Vienneau, D.; Klot, S.; Wolf, K.; Brunekreef, B.; Hoek, G., Development of Land Use Regression models for PM(2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. *Environ Sci Technol* **2012,** *46* (20), 11195-205.

Eeftens, M.; Phuleria, H. C.; Meier, R.; Aguilera, I.; Corradi, E.; Davey, M.; Ducret-Stich, R.; Fierz, M.; Gehrig, R.; Ineichen, A.; Keidel, D.; Probst-Hensch, N.; Ragettli, M. S.; Schindler, C.; Künzli, N.; Tsai, M.-Y., Spatial and temporal variability of ultrafine particles, NO2, PM2.5, PM2.5 absorbance, PM10 and PMcoarse in Swiss study areas. *Atmospheric Environment* **2015,** *111*, 60-70.

Eeftens, M.; Tsai, M.-Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; De Nazelle, A.; de Vocht, F.; Declercq, C.; Dėdelė, A.; Eriksen, K.; Galassi, C.; Gražulevičienė, R.; Grivas, G.; Heinrich, J.; Hoffmann, B.; Iakovides, M.; Ineichen, A.; Katsouyanni, K.; Korek, M.; Krämer, U.; Kuhlbusch, T.; Lanki, T.; Madsen, C.; Meliefste, K.; Mölter, A.; Mosler, G.; Nieuwenhuijsen, M.; Oldenwening, M.; Pennanen, A.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Udvardy, O.; Vaskövi, É.; Weinmayr, G.; Brunekreef, B.; Hoek, G., Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2 – Results of the ESCAPE project. *Atmospheric Environment* **2012,** *62*, 303-317.

Fang, T.; Verma, V.; Bates, J. T.; Abrams, J.; Klein, M.; Strickland, M. J.; Sarnat, S. E.; Chang, H. H.; Mulholland, J. A.; Tolbert, P. E.; Russell, A. G.; Weber, R. J., Oxidative potential of ambient water-soluble PM2.5 in the southeastern United States: contrasts in sources and health associations between ascorbic acid (AA) and dithiothreitol (DTT) assays. *Atmos. Chem. Phys.* **2016,** *16* (6), 3865-3879.

Godri, K. J.; Harrison, R. M.; Evans, T.; Baker, T.; Dunster, C.; Mudway, I. S.; Kelly, F. J., Increased oxidative burden associated with traffic component of ambient particulate matter at roadside and urban background schools sites in London. *PLoS One* **2011,** *6* (7), e21961.

Gulliver, J.; de Hoogh, K.; Hansell, A.; Vienneau, D., Development and back-extrapolation of NO2 land use regression models for historic exposure assessment in Great Britain. *Environ Sci Technol* **2013,** *47* (14), 7804-11.

Hankey, S.; Marshall, J. D., Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM2.5, Particle Size) Using Mobile Monitoring. *Environ Sci Technol* **2015,** *49* (15), 9194-202.

Henderson, S. B.; Beckerman, B.; Jerrett, M.; Brauer, M., Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ Sci Technol* **2007,** *41* (7), 2422-8.

Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D., A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* **2008,** *42* (33), 7561-7578.

Hoek, G.; Beelen, R.; Kos, G.; Dijkema, M.; van der Zee, S. C.; Fischer, P. H.; Brunekreef, B., Land use regression model for ultrafine particles in Amsterdam. *Environ Sci Technol* **2011,** *45* (2), 622-8.

Janssen, N. A.; Yang, A.; Strak, M.; Steenhof, M.; Hellack, B.; Gerlofs-Nijland, M. E.; Kuhlbusch, T.; Kelly, F.; Harrison, R.; Brunekreef, B.; Hoek, G.; Cassee, F., Oxidative potential of particulate matter collected at sites with different source characteristics. *Sci Total Environ* **2014,** *472*, 572-81.

Jedynska, A.; Hoek, G.; Wang, M.; Yang, A.; Eeftens, M.; Cyrys, J.; Keuken, M.; Ampe, C.;

Beelen, R.; Cesaroni, G.; Forastiere, F.; Cirach, M.; de Hoogh, K.; De Nazelle, A.; Nystad, W.; Akhlaghi, H. M.; Declercq, C.; Stempfelet, M.; Eriksen, K. T.; Dimakopoulou, K.; Lanki, T.; Meliefste, K.; Nieuwenhuijsen, M.; Yli-Tuomi, T.; Raaschou-Nielsen, O.; Janssen, N. A. H.; Brunekreef, B.; Kooter, I. M., Spatial variations and development of land use regression models of oxidative potential in ten European study areas. *Atmospheric Environment* **2017,** *150*, 24-32.

Kunzli, N.; Mudway, I. S.; Gotschi, T.; Shi, T.; Kelly, F. J.; Cook, S.; Burney, P.; Forsberg, B.; Gauderman, J. W.; Hazenkamp, M. E.; Heinrich, J.; Jarvis, D.; Norback, D.; Payo-Losa, F.; Poli, A.; Sunyer, J.; Borm, P. J., Comparison of oxidative properties, light absorbance, total and elemental mass concentration of ambient PM2.5 collected at 20 European sites. *Environ Health Perspect* **2006,** *114* (5), 684-90.

Liu, C.; Henderson, B. H.; Wang, D.; Yang, X.; Peng, Z. R., A land use regression application into assessing spatial variation of intra-urban fine particulate matter (PM2.5) and nitrogen dioxide (NO2) concentrations in City of Shanghai, China. *Sci Total Environ* **2016,** *565*, 607-15.

Maikawa, C. L.; Weichenthal, S.; Wheeler, A. J.; Dobbin, N. A.; Smargiassi, A.; Evans, G.; Liu, L.; Goldberg, M. S.; Pollitt, K. J., Particulate Oxidative Burden as a Predictor of Exhaled Nitric Oxide in Children with Asthma. *Environ Health Perspect* **2016,** *124* (10), 1616-1622.

Montagne, D. R.; Hoek, G.; Klompmaker, J. O.; Wang, M.; Meliefste, K.; Brunekreef, B., Land Use Regression Models for Ultrafine Particles and Black Carbon Based on Short-Term Monitoring Predict Past Spatial Variation. *Environ Sci Technol* **2015,** *49* (14), 8712-20.

Pope, C. A., 3rd; Dockery, D. W., Health effects of fine particulate air pollution: lines that connect. *J Air Waste Manag Assoc* **2006,** *56* (6), 709-42.

Rivera, M.; Basagaña, X.; Aguilera, I.; Agis, D.; Bouso, L.; Foraster, M.; Medina-Ramón, M.; Pey, J.; Künzli, N.; Hoek, G., Spatial distribution of ultrafine particles in urban settings: A land use regression model. *Atmospheric Environment* **2012,** *54*, 657-666.

Royal College of Physicians. Every breath we take: the lifelong impact of air pollution. Report of a working party. London: RCP, 2016. www.rcplondon.ac.uk/projects/outputs/every-breath-we-take-lifelong-impact-air-pollution [Accessed 10th July 2017].

Shi, T.; Duffin, R.; Borm, P. J.; Li, H.; Weishaupt, C.; Schins, R. P., Hydroxyl-radical-dependent DNA damage by ambient particulate matter from contrasting sampling locations. *Environ Res* **2006,** *101* (1), 18-24.

Tsai, M. Y.; Hoek, G.; Eeftens, M.; de Hoogh, K.; Beelen, R.; Beregszaszi, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; De Nazelle, A.; de Vocht, F.; Ducret-Stich, R.; Eriksen, K.; Galassi, C.; Grazuleviciene, R.; Grazulevicius, T.; Grivas, G.; Gryparis, A.; Heinrich, J.; Hoffmann, B.; Iakovides, M.; Keuken, M.; Kramer, U.; Kunzli, N.; Lanki, T.; Madsen, C.; Meliefste, K.; Merritt, A. S.; Molter, A.; Mosler, G.; Nieuwenhuijsen, M. J.; Pershagen, G.; Phuleria, H.; Quass, U.; Ranzi, A.; Schaffner, E.; Sokhi, R.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Taimisto, P.; Tewis, M.; Udvardy, O.; Wang, M.; Brunekreef, B., Spatial variation of PM elemental composition between and within 20 European study areas--Results of the ESCAPE project. *Environ Int* **2015,** *84*, 181-92.

van Nunen, E.; Vermeulen, R.; Tsai, M. Y.; Probst-Hensch, N.; Ineichen, A.; Davey, M.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Raffaele, D.; Ranzi, A.; Ivaldi, C.; Galassi, C.; Nieuwenhuijsen, M.; Curto, A.; Donaire-Gonzalez, D.; Cirach, M.; Chatzi, L.; Kampouri, M.; Vlaanderen, J.; Meliefste, K.; Buijtenhuijs, D.; Brunekreef, B.; Morley, D.; Vineis, P.; Gulliver, J.; Hoek, G., Land Use Regression Models for Ultrafine Particles in Six European Areas. *Environ Sci Technol* **2017,** *51* (6), 3336-3345.

Vineis, P.; Chadeau-Hyam, M.; Gmuender, H.; Gulliver, J.; Herceg, Z.; Kleinjans, J.; Kogevinas, M.; Kyrtopoulos, S.; Nieuwenhuijsen, M.; Phillips, D. H.; Probst-Hensch, N.;

Scalbert, A.; Vermeulen, R.; Wild, C. P.; Consortium, E. X., The exposome in practice: Design of the EXPOsOMICS project. *Int J Hyg Environ Health* **2016**.

Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B., Systematic evaluation of land use regression models for NO(2). *Environ Sci Technol* **2012,** *46* (8), 4481-9.

Wang, M.; Brunekreef, B.; Gehring, U.; Szpiro, A.; Hoek, G.; Beelen, R., A New Technique for Evaluating Land-use Regression Models and Their Impact on Health Effect Estimates. *Epidemiology* **2016,** *27* (1), 51-6.

Weichenthal, S.; Crouse, D. L.; Pinault, L.; Godri-Pollitt, K.; Lavigne, E.; Evans, G.; van Donkelaar, A.; Martin, R. V.; Burnett, R. T., Oxidative burden of fine particulate air pollution and risk of cause-specific mortality in the Canadian Census Health and Environment Cohort (CanCHEC). *Environmental Research* **2016,** *146*, 92-99.

Wolf, K.; Cyrys, J.; Harcinikova, T.; Gu, J.; Kusch, T.; Hampel, R.; Schneider, A.; Peters, A., Land use regression modeling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany. *Sci Total Environ* **2017,** *579*, 1531-1540.

Yang, A.; Hellack, B.; Leseman, D.; Brunekreef, B.; Kuhlbusch, T. A. J.; Cassee, F. R.; Hoek, G.; Janssen, N. A. H., Temporal and spatial variation of the metal-related oxidative potential of PM2.5 and its relation to PM2.5 mass and elemental composition. *Atmospheric Environment* **2015,** *102*, 62-69.

Yang, A.; Wang, M.; Eeftens, M.; Beelen, R.; Dons, E.; Leseman, D. L.; Brunekreef, B.; Cassee, F. R.; Janssen, N. A.; Hoek, G., Spatial Variation and Land Use Regression Modeling of the Oxidative Potential of Fine Particles. *Environ Health Perspect* **2015,** *123* (11), 1187-92.

Yanosky, J. D.; Tonne, C. C.; Beevers, S. D.; Wilkinson, P.; Kelly, F. J., Modeling exposures to the oxidative potential of PM10. *Environ Sci Technol* **2012,** *46* (14), 7612-20.

Zhang, J. J. Y.; Sun, L.; Barrett, O.; Bertazzon, S.; Underwood, F. E.; Johnson, M., Development of land-use regression models for metals associated with airborne particulate matter in a North American city. *Atmospheric Environment* **2015,** *106*, 165-177.

Figure 1. Boxplots of measured annual average concentration (% consumption) of OP^{AA} and OP^{GSH} by study area.

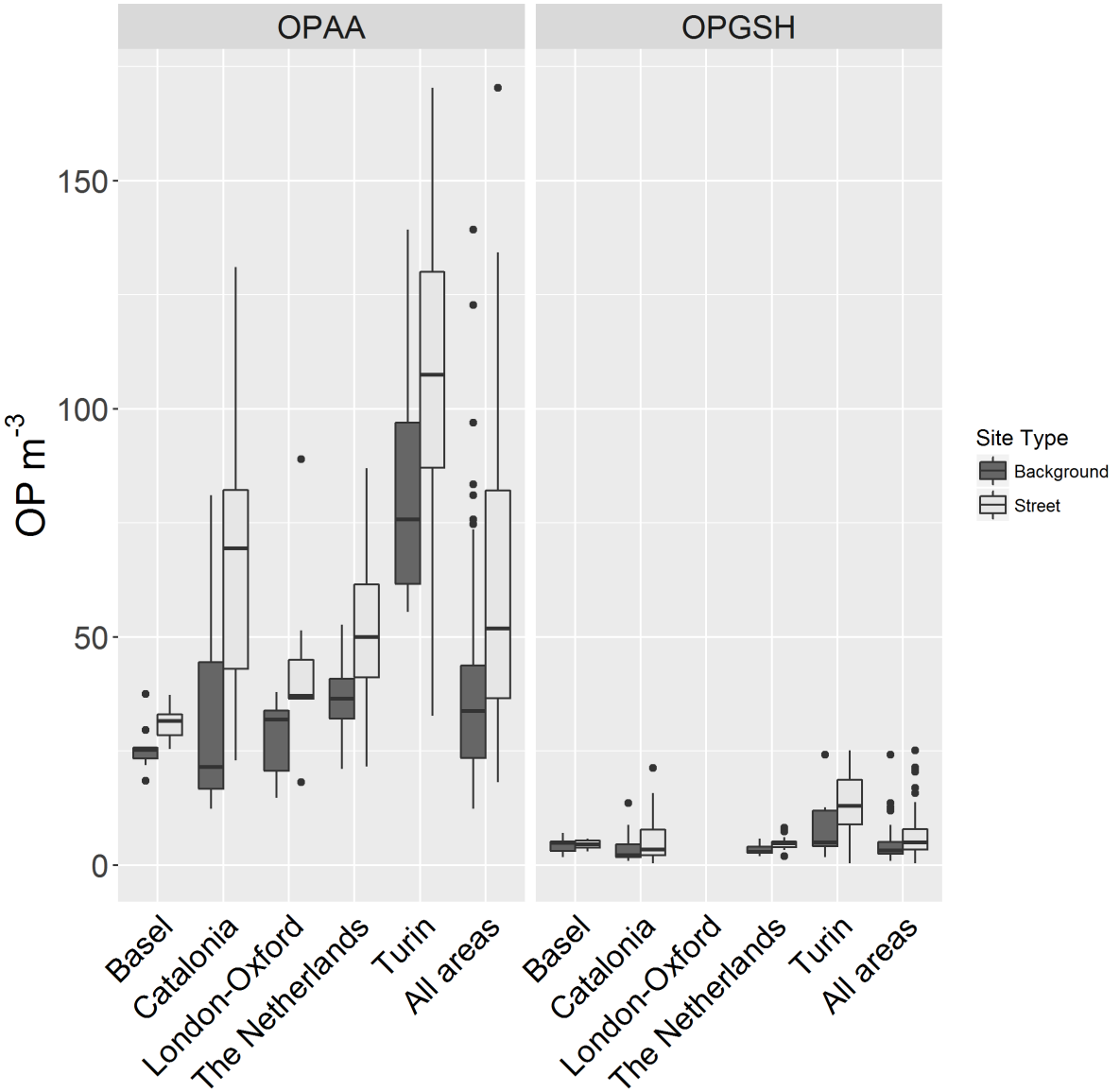


Table 1. Pearson (R) correlation of measured values of OP^{AA} (n=138) and OP^{GSH} (n=114) with measured values of $PM_{2.5}$, $PM_{2.5}$ absorbance, NO_2, Cu, Fe, K, Ni, S, Si, V, and Zn: the average (min, max) of all areas, each area individually (Basel, Catalonia, London-Oxford (only OP^{AA}), the Netherlands, and Turin), and all areas pooled.

	Basel		Catalonia		London-Oxford		The Netherlands		Turin		Average (min, max) of all areas[b]		All areas pooled	
	OP^{AA}	OP^{GSH}	OP^{AA}	OP^{GSH}	OP^{AA}	OP^{GSH}	OP^{AA}	OP^{GSH}	OP^{AA}	OP^{GSH}	OP^{AA}	OP^{GSH}	OP^{AA}	OP^{GSH}
N	20	20	39	35	20	-	39	39	20	20			138	114
OP^{GSH}	.20	-	**.67[a]**	-	-	-	**.42**		**.61**	-	.48 (.20, .61)	-	**.71[c]**	
$PM_{2.5}$	.32	.21	**.43**	.11*	**.81**	-	**.57**	**.42**	**.51**	.37*	.59 (.32, .81)	.28 (.11, .42)	**.73**	**.52**
$PM_{2.5}$absorbance	**.55**	.30	**.54**	.13*	**.84**	-	**.70**	**.41**	**.54**	.37*	.63 (.54, .84)	.30 (.13, .41)	**.72**	**.40**
NO_2	**.49**	.35	**.55**	.15*	**.88**	-	**.75**	**.39**	.43*	.43*	.62 (.43, .88)	.33 (.15, .43)	**.62**	**.37**
Cu	.35	-.17	**.45**	.14*	**.93**	-	**.71**	**.47**	**.53**	.42*	.59 (.35, .93)	.22 (-.17, .47)	**.71**	**.46**
Fe	**.74**	.05	**.52**	.25*	**.95**	-	**.72**	**.43**	**.56**	**.50**	.70 (.52, .95)	.31 (.05, .50)	**.71**	**.48**
K	-.20	.13	.29*	.13*	.18*	-	-.02*	.15*	.22*	-.10*	.09 (-.20, .29)	.08 (-.10, .13)	**.62**	**.48**
Ni	.09	-.07	.15*	-.23*	.02*	-	.32*	-.08*	.36*	**.50**	.19 (.02, .36)	.03 (-.23, .50)	**.38**	.06*
S	**.67**	.37	**.34**	-.13*	.09*	-	.23*	-.05*	**.45**	.09*	.36 (.09, .67)	.07 (-.13, .37)	**.57**	**.21**
Si	**.59**	-.39	**.36**	.12*	**.58**	-	**.45**	**.42**	.36*	.38*	.47 (.36, .59)	.13 (-.39, .42)	**.57**	**.35**
V	-.24	-.26	.02*	**-.38**	.31*	-	.29*	-.02*	.32*	.11*	.14 (-.24, .32)	-.14 (-.38, .11)	.10*	**-.22**
Zn	.37	.14	.20*	-.16*	**.67**	-	-.04*	.02*	**.54**	.32*	.35 (-.04, .67)	.08 (-.16, .32)	**.34**	.06*

Values in bold are significant at the 95% level (p > .05)
[a]N=35
[b]Average of within area correlation (hence levels of significance not applied)
[c]N=118

Table 2. Performance statistics for the OP^{AA} LUR models.

Model stage	Statistic	Basel	Catalonia	London-Oxford	The Netherlands	Turin	Combined[a]
Derivation using all sites	N (sites)	20	39	20	39	20	138
	Variables	ROADLENGTH25 LDRES5000	REFSITE TRAFLOADMAJOR100 UGNL5000	TRAFLOADMAJOR50 ROADLENGTH1000	TRAFLOAD50 POP5000	TRAFLOADMAJOR100 URBGREEN500 NATURAL1000	TRAFLOADMAJOR50 ROADLENGTH500 NATURAL5000
	R^2	0.44	0.64	0.84	0.60	0.56	0.65
	RMSE	3.77	18.76	6.53	8.82	22.42	18.35
	NRMSE	0.13	0.35	0.20	0.20	0.23	0.36
LOOCV	R^2	0.32	0.58	0.82	0.54	0.31	0.60
	RMSE	4.14	19.98	6.90	9.33	26.67	19.60
Repeated sub-sampling (RSS)[b]	N (models)	20	39	20	39	20	138
	Min R^2	0.35	0.53	0.50	0.48	0.40	0.60
	10^{th}%ile R^2	0.36	0.60	0.79	0.55	0.45	0.63
	Median R^2	0.43	0.64	0.84	0.60	0.53	0.65
	90^{th}%ile R^2	0.65	0.67	0.87	0.64	0.63	0.68
	Max R^2	0.70	0.70	0.89	0.68	0.68	0.72
	Min RMSE	2.78	17.44	5.58	7.69	19.71	16.66
	10^{th}%ile RMSE	2.98	18.08	5.99	8.32	20.82	17.48
	Median RMSE	3.86	18.80	6.70	8.85	22.65	18.32
	90^{th}%ile RMSE	3.99	19.48	6.92	9.14	23.66	18.94
	Max RMSE	3.99	19.82	6.93	9.22	23.90	19.27
	% Cook's 'D' > 1	0	0	1.9	0	5.6	0
Hold-out validation (HOV)	N (sites)	40	156	40	156	40	1918
	R^2	0.21	0.58	0.45	0.53	0.13	0.60
	RMSE	4.47	19.90	11.41	9.54	30.52	19.62
	NRMSE	0.16	0.37	0.34	0.22	0.31	0.37

[a]Combined model includes all areas using linear mixed effects to derive a model where random intercepts are used to differentiate between the effect of country.

[b]Variables from the initial models are repeatedly offered into regression analysis using a sub-set of measurement sites (N-10%) until all sites have been re-entered the maximum number of times (e.g. for 20 sites, N-10% is a maximum of 2 entries per site yielding 20 models).

Variable names followed by values of radii (m) of circular buffers: LDRES – low density residential land; NATURAL – semi-natural and forested areas; POP – number of inhabitants; REFSITE – ID of reference site (Catalonia had three reference sites); ROADLENGTH – length of all roads; TRAFLOAD – traffic load on all roads; TRAFLOADMAJOR – traffic load on all major roads; UGNL – sum of URBGREEN and NATURAL; URBGREEN – urban green space.

Table 3. Performance statistics for the OP^{GSH} LUR models.

Model stage	Statistic	Basel	Catalonia	London-Oxford	The Netherlands	Turin	Combined[a]
Derivation using all sites	N (sites)	20	35	-	39	20	114
	Variables		REFSITE		INTMAJORINVDIST	TRAFLOAD1000	TRAFLOADMAJOR50
			TRAFLOADMAJOR100		NATURAL5000	INDUSTRY1000	
	R^2	-	0.51	-	0.44	0.22	0.39
	RMSE	-	3.35	-	1.07	6.61	3.80
	NRMSE	-	0.67	-	0.27	0.60	0.68
LOOCV	R^2	-	0.38	-	0.35	0.05	0.28
	RMSE	-	3.74	-	1.14	7.61	4.13
Repeated sub-sampling (RSS)[b]	N (models)	-	*	-	39	20	114
	Min R^2	-	*	-	0.23	0.12	0.25
	10^{th}%ile R^2	-	*	-	0.38	0.13	0.33
	Median R^2	-	*	-	0.44	0.20	0.36
	90^{th}%ile R^2	-	*	-	0.49	0.29	0.40
	Max R^2	-	*	-	0.59	0.32	0.46
	Min RMSE	-	*	-	0.92	6.14	3.34
	10^{th}%ile RMSE	-	*	-	1.02	6.22	3.65
	Median RMSE	-	*	-	1.09	6.54	3.91
	90^{th}%ile RMSE	-	*	-	1.12	6.62	4.03
	Max RMSE	-	*	-	1.13	6.64	4.06
	% Cook's 'D' > 1	-	*	-	0.3	1.1	0
Hold-out validation (HOV)	N (sites)	-	*	-	156	40	1243
	R^2	-	*	-	0.31	-0.03	0.27
	RMSE	-	*	-	1.17	7.85	4.14
	NRMSE	-	*	-	0.29	0.71	0.74

[a]Combined model includes all areas using linear mixed effects to derive a model where random intercepts are used to differentiate between the effect of country.
[b]Variables from the initial models are repeatedly offered into regression analysis using a sub-set of measurement sites (N-10%) until all sites have been re-entered the maximum number of times (e.g. for 20 sites, N-10% is a maximum of 2 entries per site yielding 20 models).
*Statistically significant model could not be derived.
INTMAJORINVDIST – product of inverse distance to- and traffic intensity on- nearest major road. Variable names followed by values of radii (m) of circular buffers: NATURAL – semi-natural and forested areas; INDUSTRY – area of industrial land; REFSITE – ID of reference site (Catalonia had three reference sites); TRAFLOAD – traffic load on all roads; TRAFLOADMAJOR – traffic load on all major roads.

SUPPORTING INFORMATION

Table of Contents:

Table S1. Summary of PM2.5 filter collections

Filter collections						
		SEASON				
			Winter	Spring	Summer	Autumn
Country	City	Site Type[a]	Filter numbers (+ field blanks)			
THE NETHERLANDS	-	RB	10 (3)	17 (2)	10 (4)	13 (2)
		S	14	17	9	9
		UB	6	12	6	11
ITALY	Turin	RB	0	1	0	2
		S	8	6	10	9
		UB	5	14	11	13
		unknown	(1)	(1)	(2)	(1)
SPAIN	Catalonia	RB	5	10	8	9
		S	16	23	11	14
		UB	20	27	18	21
		unknown	(1)	(1)	(1)	(2)
UNITED KINGDOM	London-Oxford	RB	1	1	1	0
		S	5	1	8	6
		UB	15 (2)	7 (1)	21 (2)	7 (1)
SWITZERLAND	Basel[b]	RB	1	2	-	-
		S	17	15	-	-
		UB	14	15	-	-
		unknown	1	-	-	-

[a]RB = regional background, S= suburban, UB = urban background

[b]**Field blanks were collected in several regions in Switzerland as part of the SAPALDIA study but none were collected in Basel. An average value for the limit of detection for $PM_{2.5}$ was determined from the series of field blanks and this was assumed to be representative of Basel (Eeftens et al., 2015). None of the filters were below the detection limit.**

SA.1 Extraction of $PM_{2.5}$ from Teflon (PTFE) filters with PMP ring

- The Teflon filters were stored at -20^{o}C. Before the extraction was undertaken the filters were removed and allowed to condition to room temperature before opening filter holder. This was undertaken to prevent condensation of moisture onto the cold filter.
- Teflon lab blanks* and field blanks** were included in the following procedure.
- The Teflon filters, at room-temperature, were transferred with clean forceps to well-labelled small clean petri-dishes (Pall Life Sciences 7232, USA), and extracted in batches of 30.
- The sonicating water-bath (Nickel-Electro Ltd., UK, model SW12H) was set to 40^{o}C.
- 1mL of HPLC-grade methanol was added to the filter surface, the lid replaced, and the petri-dish placed securely in the water-bath.
- The sonication was activated for 30 seconds on “boost” mode.
- The petri-dish was removed from the water-bath, the excess water removed with tissue, and the petri-dish lid carefully removed. The filter was flipped over (using very clean forceps), and the lid replaced.
- The petri-dish was placed securely back into the water-bath and sonicated for another 5 minutes, initially for 30 seconds in “boost” mode followed by 4.5 minutes in “Sweep action” mode.
- On completion of the sonication time, the petri-dish was removed from the water-bath, the excess water removed with tissue, and the petri-dish lid carefully removed.
- The methanol in the petri-dish was carefully transferred to pre-weighed 2mL tubes*** (Alpha Labs., UK.), and the tube capped.
- Another 1mL methanol was added to the filter (still face down) and the 5-minute sonication procedure was repeated. This methanol extract was placed into the same tube as the 1st methanol extract.
- The combined methanol extracts were then dried down under nitrogen gas with gentle warming (37°C) using a N_2 evaporator/water bath in a fume cupboard.
- The tube containing the dried-down extracted $PM_{2.5}$ was post-weighed.
- The pre- and post-weights of the tube was used to calculate the final extracted $PM_{2.5}$ mass.
- The extracted filter was flipped back over (with clean forceps) to its original face-up position and allowed to dry (in petri-dish with lid slightly ajar) in a fume cupboard.
- Once the filter was completely dry, the petri-dish was closed and archived for further evaluation if required.

* Lab blanks were of the same matrix/size/type as the experimental filters that had never left the Lab or its filter cartridge holder before processing.
**Field blanks were of same batch number as the experimental filters, had travelled with the experimental filters out to the same field/site, removed from filter cartridge holder and returned to filter cartridge holder under same conditions.
*** the 2mL tubes were weighed BEFORE and AFTER loading with the $PM_{2.5}$ extract to determine the $PM_{2.5}$ mass that had been extracted. A QA/QC weighing procedure was

followed and was undertaken in-house within a controlled environment (50% humidity, 21^{o}C temperature).

Re-suspension of extracted $PM_{2.5}$

- The calculated $PM_{2.5}$ mass extracted was used to calculate the amount of re-suspension liquid (ultra-pure Chelex100-treated H_2O, pH7.0, containing 5% v/v methanol) to add. This would provide a stock particulate concentration of between 150 and 500µg/mL (depending on the amount of $PM_{2.5}$ extracted).
- The samples were vigorously mixed for 10 minutes, and the $PM_{2.5}$ still visibly coating the sides of the tube were gently scraped off.
- The $PM_{2.5}$ suspensions were placed on ice, along with a probe blank*, lab blank and field blank** tubes.
- The $PM_{2.5}$ suspensions, whilst the tubes were kept on ice, were then directly sonicated (MSE Soniprobe, UK) with the probe set at an amplitude of 10 microns for 30 seconds.
- In-between each PM sonication, the sonication probe was thoroughly cleaned by immersion and sonication in chelex100-treated water pH7.0.
- The re-suspended $PM_{2.5}$ were stored in 0.5 - 1mL aliquots (1.5mL micro-centrifuge tubes with an O ring seal (T332-5, Simport, Canada) and frozen to -70^{o}C until required.

***** The *probe blank* tube was treated in the same way as the PM suspension tubes *i.e*. the same amount of 5% v/v MeOH in chelex-treated water was added to the tube, probe-sonicated and frozen in aliquots. ****** The *lab and field blanks* tube were treated in the same way as the PM suspension tube; *i.e*. the same amount of 5% v /v MeOH in chelex-treated water, pH7.0, was added to the tube, probe-sonicated and frozen in aliquots and stored with its appropriate country/site batch.

SA.2 Antioxidant analysis methods

Determination of glutathione (GSH)

This assay employs the technique of the GSSG reductase-DTNB linked assay.[1] An aliquot (16.7µL) of the centrifuged RTLF-exposed liquid was added to 983.3µL of cold 100mmoles/L Sodium phosphate buffer, pH7.5, containing 1mmoles/L EDTA, mixed and stored at -70^{o}C. On the day of analysis the sample was thawed, placed on ice immediately it was thawed, and 50µL analysed (in duplicate) in parallel with glutathione standards for total both (GSX) glutathione and (following derivitization with 2-vinyl pyridine) for oxidised (GSSG) glutathione. The reduced (GSH) glutathione was obtained by subtraction of the GSSG value from the GSX. The %CV of analysis was less than 10% with a minimum detection limit of 0.3µmoles/L.

The microplate reader used was a Spectramax190 (Molecular Devices, UK) along with the SoftMaxPro v4.8 software.

Determination of Ascorbic acid (AA)

This assay employs the technique of high performance liquid chromatography (HPLC) with an electrochemical detector[2] with modifications. 50µL of the centrifuged RTLF-exposed liquid was added to 450µL of cold 5.6% v/v *meta*-phosphoric acid in 0.7ml amber HPLC vial, mixed, and either stored at -70^{o}C or immediately analysed on the HPLC. Aliquots of 20 µL acidified sample were injected onto a 150 x 4.6mm 5µ SphereClone ODS(2) column (Phenomenex, UK) and eluted with a 0.2 mol/L K_2HPO_4-H_3PO_4 (pH 2.1) mobile phase containing 0.25 mmol/L octanesulfonic acid.

Final concentrations for ascorbic (AA) acid was calculated with external AA standards, which were run simultaneously. The %CV of analysis was less than 5% with a minimum detection limit for ascorbic acid of 0.5µmol/L and uric acid of 0.1µmol/L.

The HPLC used was purchased from Gilson Scientific UK along with the Unipoint v5.1 software

All chemicals were of the highest grade possible, usually HPLC-grade, and purchased from either the Sigma Chemical Company (UK) or VWR (UK) or Fisher Scientific (UK).

References:

1. Baker MA, Cerniglia GJ, Zaman A (1990) *Anal Biochem* **190**: 360–365.
2. IIriyama, K., Yoshiura M., Iwamoto T. and Ozaki Y. 1984. *Anal. Biochem.* **141**:238-243.

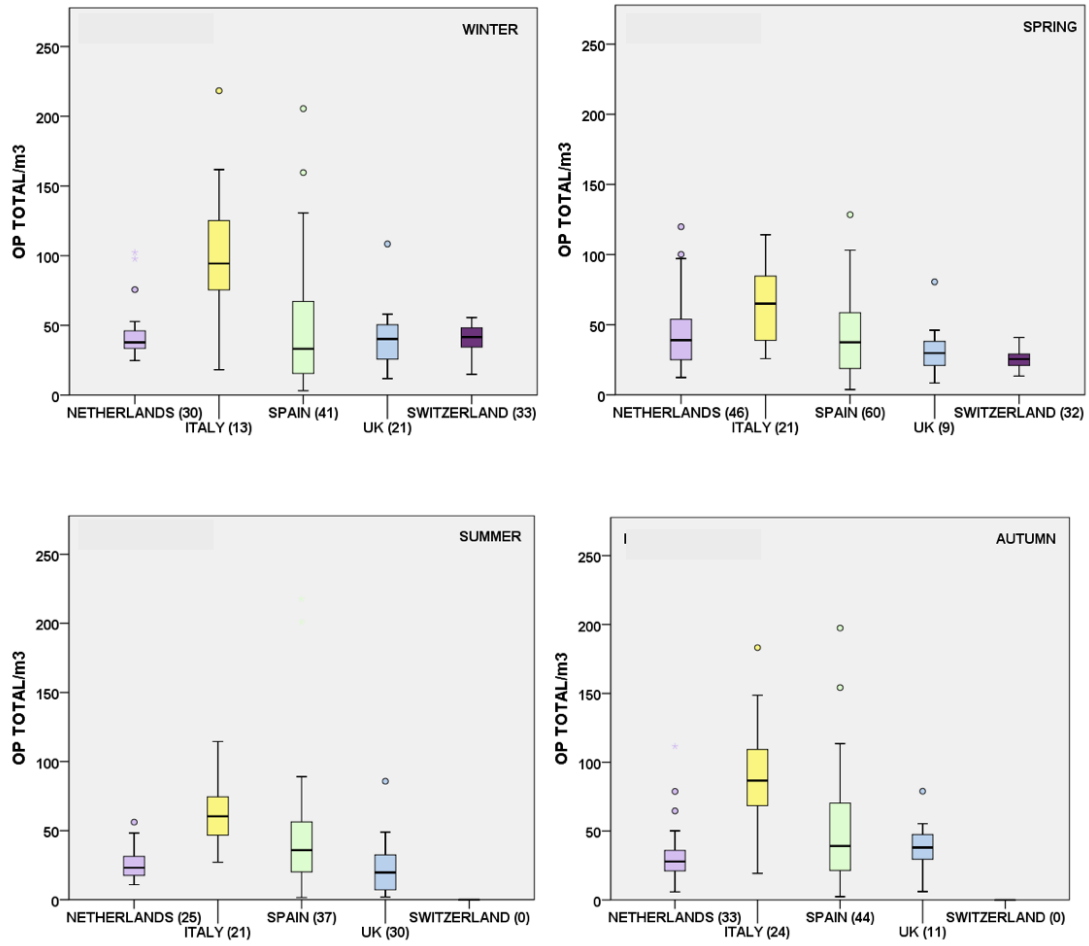
SA.3 RTLF model - in-house Controls

A starting concentration of 200µmol/L antioxidants (C0) in the synthetic respiratory tract fluid (RTLF) displayed a minimal amount of oxidation after an incubation of 4 hours (C4). The remaining antioxidant concentrations in the RTLF after incubation with the in-house controls were as expected for the RTLF model (see table below). The –ve control PM (M120) displayed no reactivity with the antioxidant, whereas the +ve control PM (NIST1648a) displayed approx. 50% consumption of the AA. The lab filter blank of a similar Teflon matrix displayed no reactivity with the antioxidants, as did the probe blank.

µmole/L ANTIOXIDANT REMAINING AFTER 4 HOUR INCUBATION of IN-HOUSE CONTROLS								
	AA		GSX		GSSG		GSH	
	X	±SD	**X**	±SD	**X**	±SD	**X**	±SD
C0	**200.0**	5.5	**200.00**	4.0	**0.0**	0.7	**200.0**	3.3
C4	**189.9**	4.4	**198.4**	3.5	**7.0**	1.9	**187.8**	4.5
M120	**182.0**	16.1	**202.1**	8.1	**8.1**	4.2	**185.9**	11.9
NIST1648a	**58.8**	11.7	**199.2**	7.2	**12.7**	13.1	**173.8**	25.1
Lab filter blank	**184.6**	11.0	**214.0**	8.9	**7.1**	3.1	**200.0**	10.8
Probe blank	**184.4**	5.1	**200.0**	4.7	**5.0**	4.2	**190.0**	9.7

SA.4. Establishing OP mass and volume

The oxidative consumption of the antioxidants ascorbic acid (AA) and glutathione (GSH) in the presence of the $PM_{2.5}$ was measured and converted to the % consumed; $\%OP^{AA}$ and $\%OP^{GSH}$, by reference to the oxidation measured for in-house particle-free 4-hour Control.
The $\%OP^{AA}$ and $\%OP^{GSH}$ data were converted to OP^{AA}/µg $PM_{2.5}$ and OP^{GSH}/µg $PM_{2.5}$ based on the known $PM_{2.5}$ mass present in the RTLF exposure tube. The %OP data had also been previously corrected for the background oxidation observed for its particular site field blank (field blanks were not available for Basel).
The Oxidative Potential of the total mass of $PM_{2.5}$ present per volume of air (m^3), an indicator of the toxicity of the surrounding environment, was calculated for all countries both for sites and seasons. By multiplying the OP/µg $PM_{2.5}$ data by the $PM_{2.5}/m^3$ data, the oxidative consumption of AA and GSH per m^3, could be reported as OP^{AA}/m^3 and OP^{GSH}/m^3.
The OP^{TOTAL}/m^3 data for each of the five areas [Basel (Switzerland); Catalonia (Spain); the Netherlands; London-Oxford (UK), Turn (Italy)] is shown below for the four seasons (Winter, Spring, Summer, and Autumn).



Variation of TOTAL oxidative potential per unit volume (OP^{TOTAL} m^{-3}) measured for each season; Winter, Spring, Summer, and Autumn, for all participating countries $PM_{2.5}$. The box & whisker plot format indicates the interquartile range (the box), the median (line in the box), the minimum and maximum range (whiskers) and outliers (stars/circles).

SA.5. Calculation of 'annual' average values of oxidative potential

Each study area had a continuous reference site as the basis for temporal adjustment (except in Catalonia there were 3 sites; 1 per each distinct sub-area). For temporal adjustment, values of OP^{AA} m^{-3} and OP^{GSH} m^{-3} at each site were adjusted by the difference between OP^{AA} m^{-3} and OP^{GSH} m^{-3} at the reference site for the corresponding period and the average of all values of OP^{AA} m^{-3} and OP^{GSH} m^{-3}at the reference site. Missing data at the reference site for London-Oxford (22 out of 59 two-week measurements), was imputed with the relationship of $PM_{2.5}$ and OP^{AA} (r^2 = 0.75) across all other sites. A weak relationship of $PM_{2.5}$ and OP^{GSH} across all sites meant that average OP^{GSH} were not calculated and hence LUR models were not developed for London-Oxford. Annual averages were calculated for each site where there were at least two values of OP^{AA} or OP^{GSH}. OP^{TOTAL} was also calculated as the sum of OP^{AA} and OP^{GSH} where data for both was available.

Table S2. Summary of OP measurements by centre and the number of measurements available for LUR modelling following data screening and temporal adjustment.

	Basel		Catalonia		Turin		The Netherlands		London-Oxford	
	OPAA	OPGSH	OPAA	OPGSH	OPAA	OPGSH	OPAA	OPGSH	OPAA	OPGSH
sites	24		40		20		40		20	
two-week measurements	59		116		60		118		59	
two-week measurements following temporal adjustment	59	58	111	96	59	54	111	109	57	-
sites with annual average	20	20	39	35	20	20	39	39	20	-

The requirement for annual average OP was valid filter measurements following temporal adjustment for two or more two-week periods representing different seasons. The main source of missing filters was either due to a lack of reference site data (and a relationship too weak between OP and PM2.5 to impute reference site data) or negative values following temporal adjustment. Details for specific areas:
Basel - 4 sites with only 1 filter; Catalonia - two periods with missing reference data relating to 14 filters; this left one site with only one filer for OPAA (n=39) and a further four sites for OPGSH had only one filter following negative values in temporal adjustment (n=35); Turin - all sites had at least two filters following temporal adjustment; The Netherlands - one site removed due to having only a single filter to represent an annual average; London-Oxford - all sites for OPAA had at least two filters following temporal adjustment but there was no data for OPGSH.

Table S3. GIS predictor variables offered in LUR model development.

Predictor Variable	Variable Name	Units	Direction	Buffer sizes (m)
Corine land use predictors				
Industry	INDUSTRY	m^2	+	100, 300, 500, 1000, 5000
Port	PORT	m^2	+	100, 300, 500, 1000, 5000
Airport	AIRPORT	m^2	+	1000, 5000
Urban Green	URBGREEN	m^2	-	100, 300, 500, 1000, 5000
Semi-natural and forested areas	NATURAL	m^2	-	100, 300, 500, 1000, 5000
Low density residential land	LDRES	m^2	+	100, 300, 500, 1000, 5000
High density residential land	HDRES	m^2	+	100, 300, 500, 1000, 5000
Sum of low and high density residential land	HDLDRES	m^2	+	100, 300, 500, 1000, 5000
Sum of URBGREEN & NATURAL	UGNL	m^2	-	100, 300, 500, 1000, 5000
Other spatial predictors (using best available local data)				
Population data	POPEEA	m^2	+	100, 300, 500, 1000, 5000
Household density	HHOLD	Number	+	100, 300, 500, 1000, 5000
Traffic intensity on nearest road	TRAFNEAR	Veh. day^{-1}	+	
Inverse distance to nearest road	DISTINVNEAR1	m^{-1}	+	
Product of traffic intensity on nearest road and inverse distance to the nearest road	INTINVDIST	Veh. $day^{-1}m^{-1}$	+	
Traffic intensity on nearest major road	TRAFMAJOR	Veh. day^{-1}	+	

Inverse distance to nearest major road	DISTINVMAJOR1	m^{-1}	+	
Product of traffic intensity in nearest major road and inverse of distance to nearest major road	INTMAJORINVDIST	Veh. $day^{-1}m^{-1}$	+	
Total traffic load of major roads in a buffer (sum of(traffic intensity*length of all segments))	TRAFLOADMAJOR	Veh. $day^{-1}m$	+	50, 100, 300, 500, 1000
Traffic total load of roads in a buffer (sum of(traffic intensity * length of all segments))	TRAFLOAD	Veh. $day^{-1}m$	+	50, 100, 300, 500, 1000
Heavy-duty traffic intensity om nearest road	HEAVYTRAFNEAR	Veh. day^{-1}	+	
Product of heavy-duty traffic intensity on nearest road and inverse distance to the nearest road	HEAVYINTINVDIST	Veh. $day^{-1}m^{-1}$	+	
Heavy-duty traffic intensity om nearest major road	HEAVYTRAFMAJOR	Veh. day^{-1}	+	
Total heavy-duty traffic load of all major roads in a buffer (sum of(heavy-duty traffic intensity * length of all segments)	HEAVYTRAFMAJORLOAD	Veh. $day^{-1}m$	+	50, 100, 300, 500, 1000
Total heavy-duty traffic load of all roads in a buffer (sum of(heavy-duty traffic intensity * length of all segments)	HEAVYTRAFLOAD	Veh. $day^{-1}m$	+	50, 100, 300, 500, 1000
Road length of all roads in a buffer	ROADLENGTH	m	+	50, 100, 300, 500, 1000
Road length of all major roads in a buffer	MAJORROADLENGTH	m	+	50, 100, 300, 500, 1000
Square root of altitude above sea level	SQRALT	m	-	

Table S4. All sites models for OP^{AA}.

Study area	N	Variable	Buffer (m)	Constant	β	Incremental adjusted R^2	Decreasing RMSE	Sig. (p)	VIF
Basel	20			6.49					
		Length of all roads	25		$1.005x10^{-1}$	0.257	4.36	0.012	1.02
		Low density residential land	5000		$5.018x10^{-7}$	0.443	3.77	0.017	1.02
Catalonia	39			-					
		Site specific factor (ref. site 82)	-		6.986				
		Site specific factor (ref. site 83)	-		$4.679x10^{1}$	0.470	22.74	0.000	1.13
		Traffic load on major roads	100		$1.428x10^{-6}$	0.561	20.69	0.009	1.07
		Urban green space, semi-natural land and forest	5000		$(-)9.875x10^{-7}$	0.639	18.76	0.006	1.22
London-Oxford	20			8.78					
		Traffic load on major roads	50		$5.062x10^{-6}$	0.722	8.52	0.000	1.68
		Length of all roads	1000		$4.596x10^{-4}$	0.836	6.53	0.002	1.68
The Netherlands	39			29.87					
		Traffic load on all roads	50		$5.605x10^{-6}$	0.380	10.99	0.000	1.06
		Number of inhabitants	5000		$4.841x10^{-5}$	0.601	8.82	0.000	1.06
Turin	20			99.04					
		Traffic load on major roads	100		$3.223x10^{-6}$	0.330	27.79	0.009	1.08
		Urban green space	50		$(-)3.305x10^{-4}$	0.422	25.83	0.018	1.08
		Semi-natural land and forest	1000		$(-)5.628x10^{-5}$	0.564	22.42	0.021	1.08
Combined[a]	138			31.31					
		Traffic load on major roads	50		$5.57x10^{-6}$	0.167	28.47	0.000	1.08
		Length of all roads	500		$1.194x10^{-3}$	0.194	28.00	0.015	1.07
		Semi-natural land and forest	5000		$(-)4.908x10^{-7}$	0.211	27.71	0.053	1.00
		Study area[b]	-			0.650	18.35		

[a]combined models include all 5 study areas using linear mixed effects.
[b]variable for 'study area' with random slopes and intercepts hence no fixed values of constant, β, p, and VIF.

Table S5. All sites models for OP^{GSH}.

Study area	N	Variable	Buffer (m)	Constant	β	Incremental adjusted R^2	Decreasing RMSE	Sig. (p)	VIF
Basel	20	No Model Derived							
Catalonia	35			-					
		Site specific factors (ref. site 82)	-		(-) 6.598	0.446	3.57	0.000	1.11
		Traffic load on major roads	100		2.163×10^{-7}	0.513	3.35	0.025	1.11
London-Oxford	-								
The Netherlands	39			3.90					
		Product of inverse distance to- and traffic intensity on- nearest major road	-		5.368×10^{-4}	0.379	1.12	0.000	1.00
		Semi-natural land and forest	5000		$(-)7.028 \times 10^{-8}$	0.435	1.07	0.038	1.00
Turin	20			1.44					
		Traffic load on all roads	1000		3.975×10^{-8}	0.065	7.22	0.032	1.16
		Industrial land	1000		1.549×10^{-5}	0.218	6.61	0.048	1.16
Combined[a]	114			4.98					
		Traffic load on major roads	50		6.646×10^{-7}	0.071	4.69	0.002	1.00
		Study area[b]	-		-	0.390	3.80	-	-

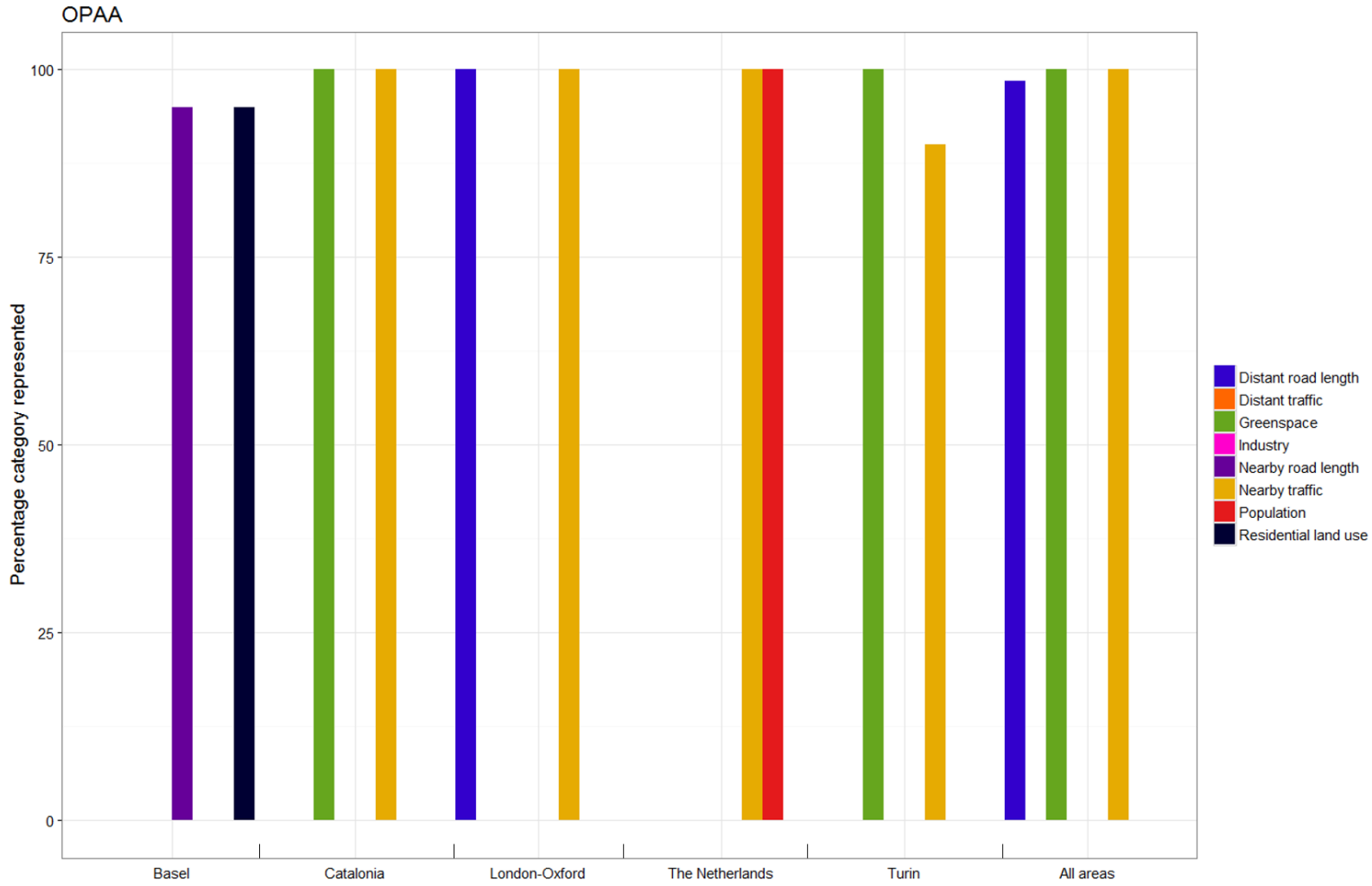
[a]combined models include all 5 study areas using linear mixed effects.
[b]variable for 'study area' with random slopes and intercepts hence no fixed values of constant, β, p, and VIF.

Table S6. Ratios (using mean values by site type) of regional background (RB), street (S) and urban background (UB) sites for OP^{AA}, OP^{GSH}, $PM_{2.5}$, $PM_{2.5}$absorbance, NO_2, Cu, Fe, K, Ni, S, Si V, and Zn [values of S/RB and UB/RB not shown for Basel, Catalonia, London-Oxford, and Turin due to the low number RB sites in those areas (N=1 except Catalonia where N=4)].

Component	All areas pooled			Basel	Catalonia	London-Oxford	The Netherlands			Turin
	S/UB	S/RB	UB/RB	S/UB	S/UB	S/UB	S/UB	S/RB	UB/RB	S/UB
OP^{AA}	1.5**	2.0**	1.4	1.2*	1.7**	1.5*	1.3**	1.6**	1.2	1.2
OP^{GSH}	1.4	2.2*	1.6	1.0	1.4		1.5**	1.3*	0.9	1.5
$PM_{2.5}$	1.2**	1.3*	1.0	1.1	1.3**	1.3*	1.1**	1.2**	1.0	1.1*
$PM_{2.5}$absorbance	1.6**	2.1**	1.3**	1.1	1.6**	1.9**	1.5**	1.9**	1.2*	1.4**
NO_2	1.6**	2.8**	1.8**	1.1	1.6**	1.7*	1.4**	2.1**	1.5**	1.8**
Cu	1.8**	3.8**	2.1**	1.2	1.8**	1.6	1.8**	2.5**	1.4*	1.9**
Fe	1.8**	3.1**	1.7**	1.3	1.8**	1.8*	1.8**	2.5**	1.4*	1.8**
K	1.1	1.4*	1.3	1.1	1.1	1.0	1.1	1.1	1.0	1.0
Ni	1.1	1.5	1.3	0.8	0.9	0.8	0.9	1.3	1.3	1.3*
S	1.1	1.1*	1.1	1.1	1.0	0.9	1.0	1.1	1.1	1.1
Si	1.4**	1.7**	1.1	1.2	1.2*	1.2	1.7**	1.6*	0.9	1.4**
V	1.0	1.4	1.3	0.8	0.9	0.9	0.9	1.2	1.3	1.1
Zn	1.3*	1.5*	1.2	1.2**	1.3	1.1	1.1	1.1	1.1	1.3**

**p<0.01. *p<0.05.

Figure S1.The number of times each variable was used in repeated sub-sampling.



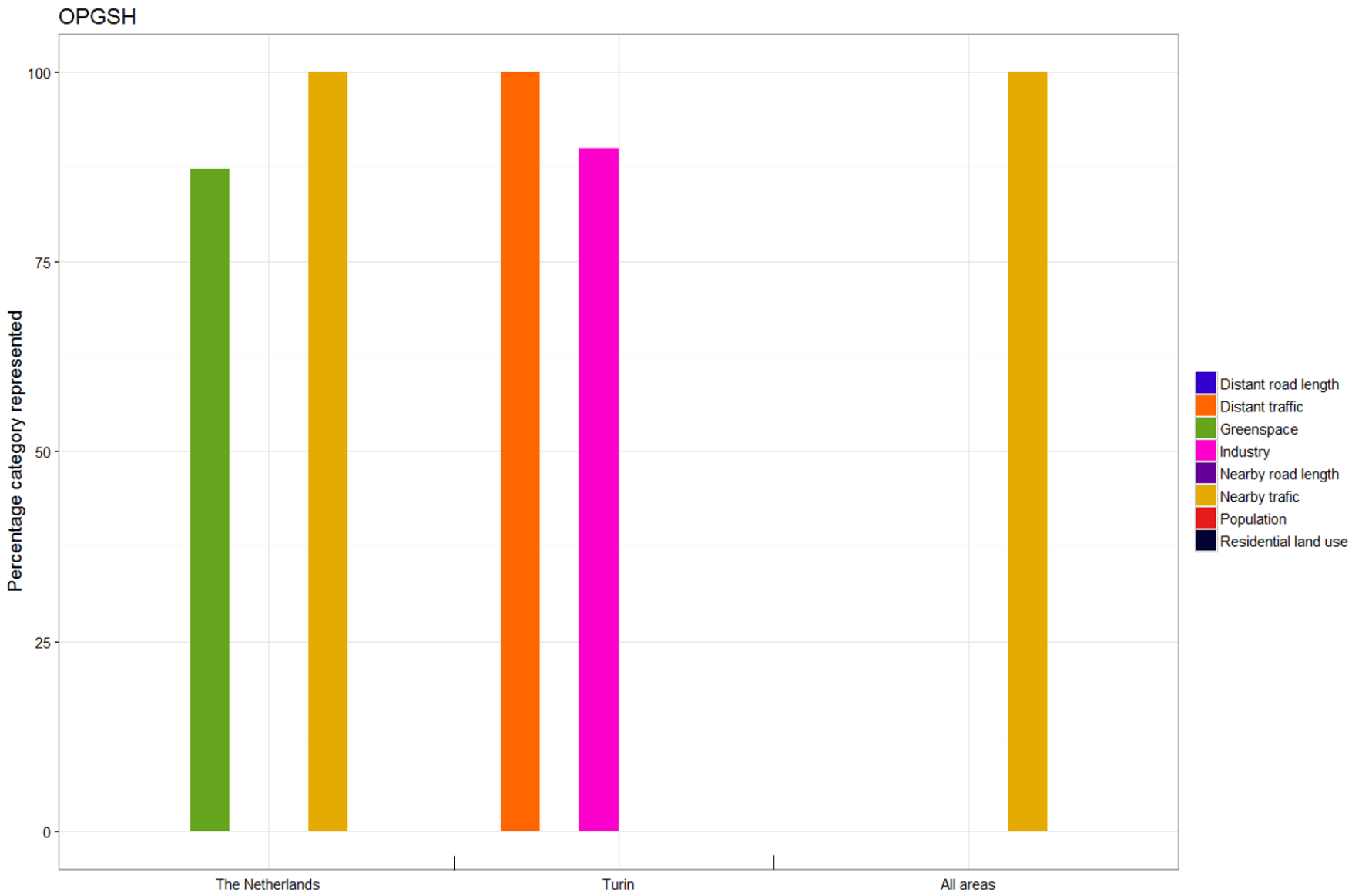
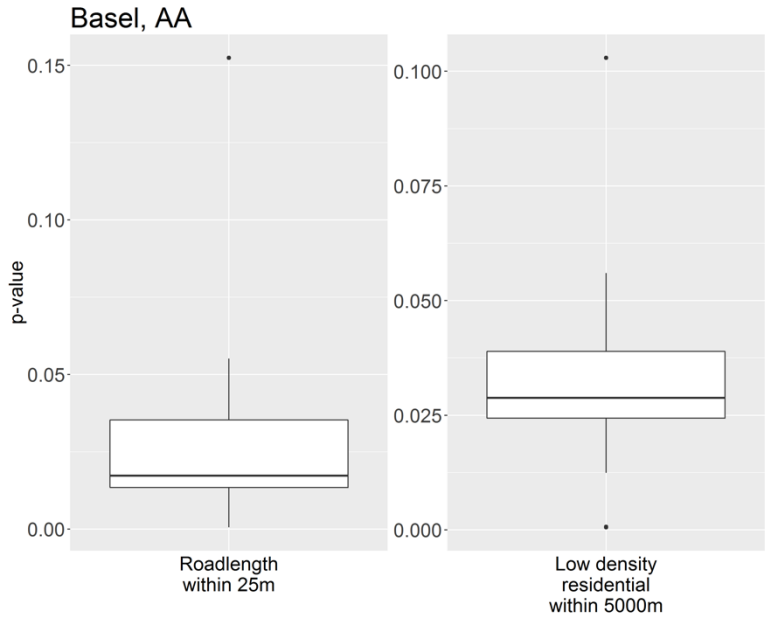

OPGSH
Percentage category represented
100
75
50
25
0
The Netherlands
Turin
All areas
Distant road length
Distant traffic
Greenspace
Industry
Nearby road length
Nearby trafic
Population
Residential land use

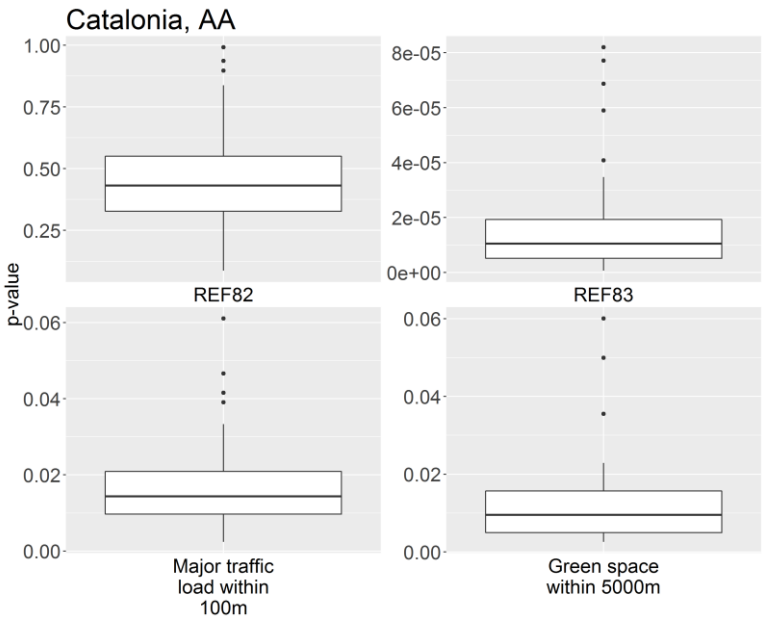
Figure S2. Boxplots of variability in individual variable p-values from RSS by study area.

Basel



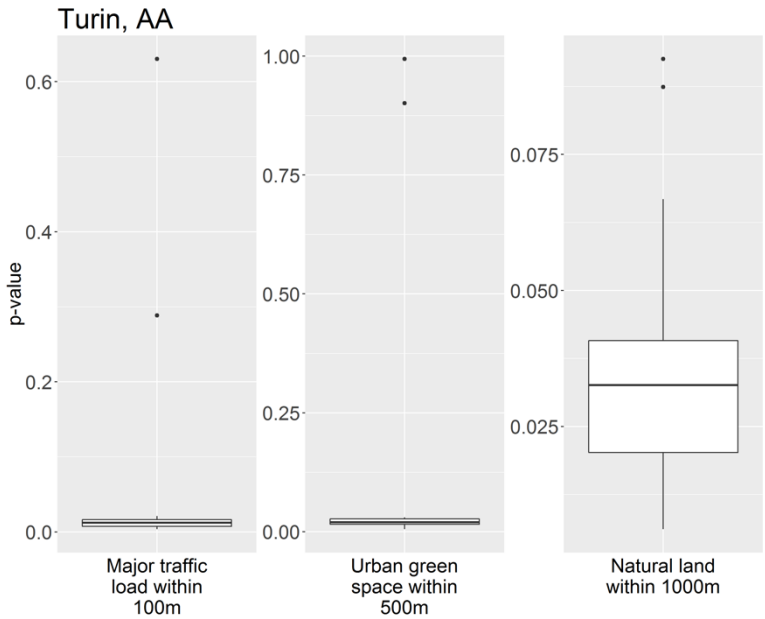
GSH – no model

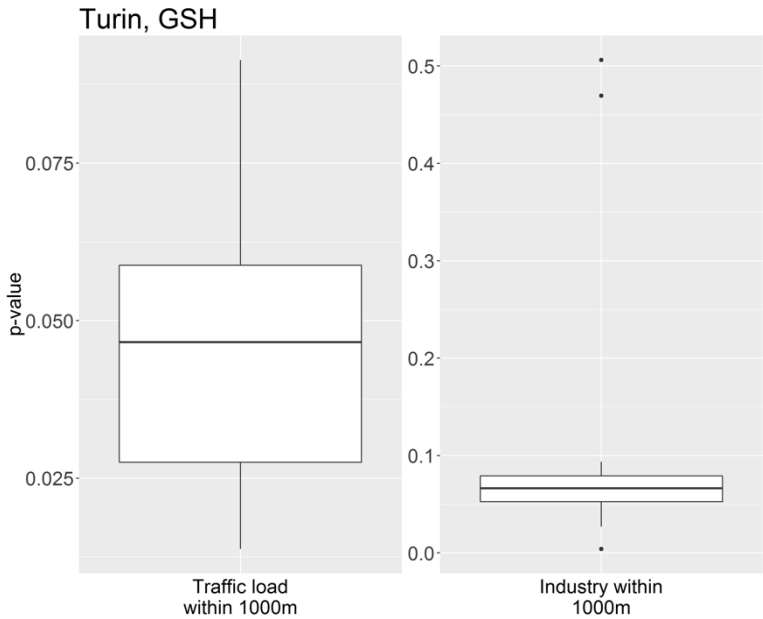
Catalonia



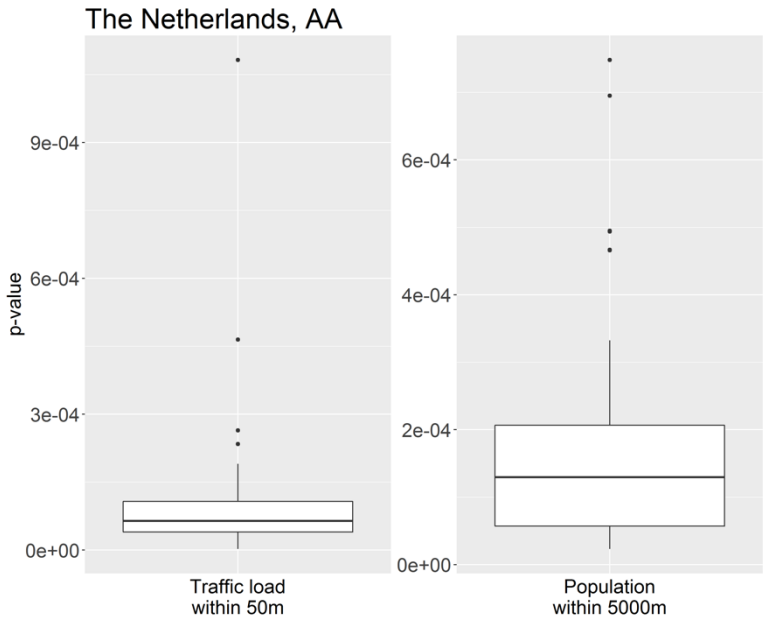
GSH – no model

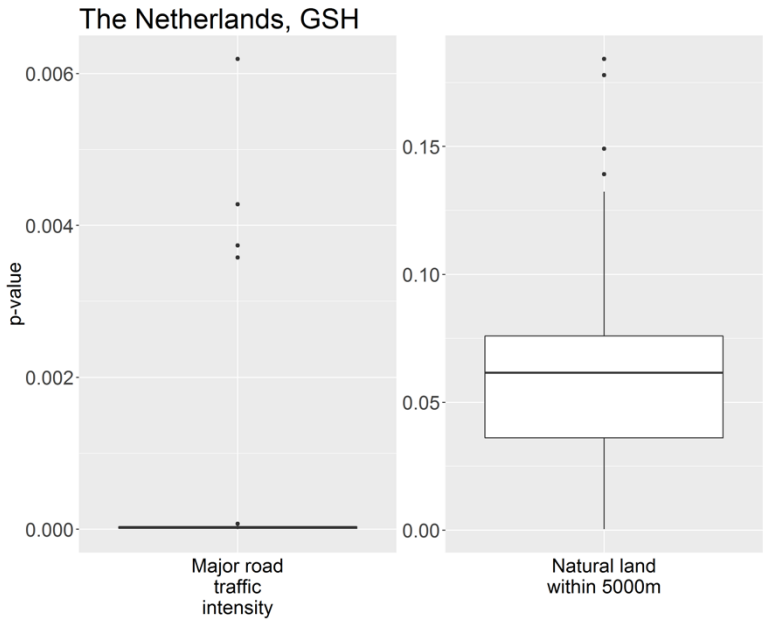
Turin


Turin, AA
p-value
0.6
0.4
0.2
0.0
1.00
0.75
0.50
0.25
0.00
0.075
0.050
0.025
Major traffic load within 100m
Urban green space within 500m
Natural land within 1000m

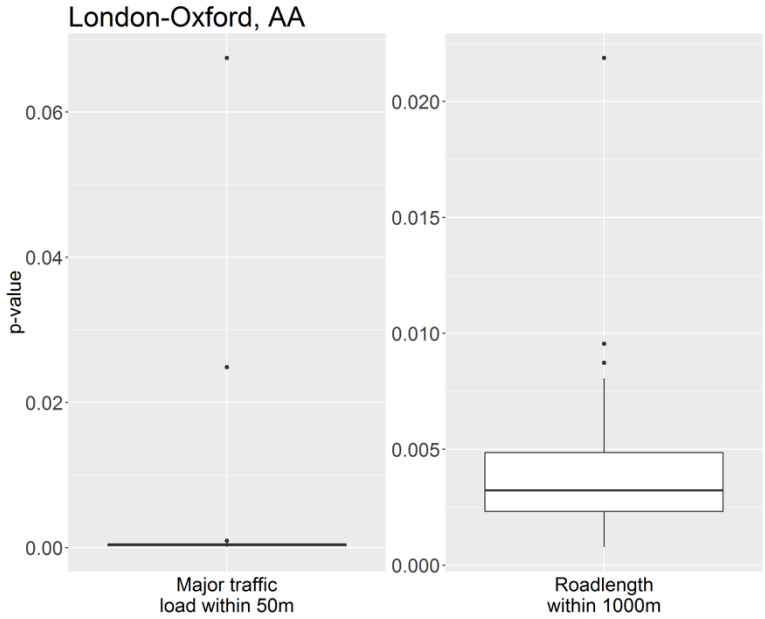

Turin, GSH
p-value
0.075
0.050
0.025
0.5
0.4
0.3
0.2
0.1
0.0
Traffic load within 1000m
Industry within 1000m

The Netherlands



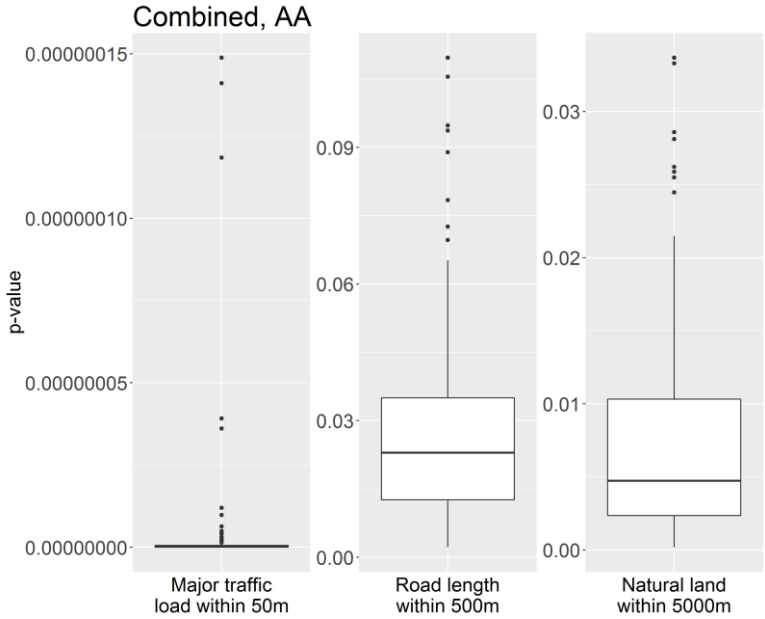


London-Oxford



GSH – no model

Combined models (all areas)



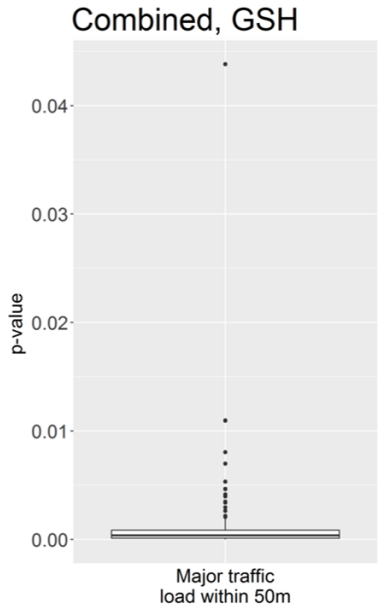
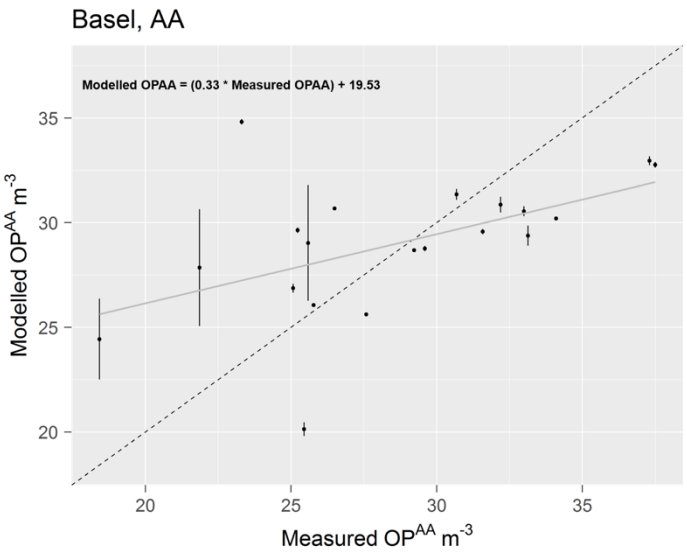


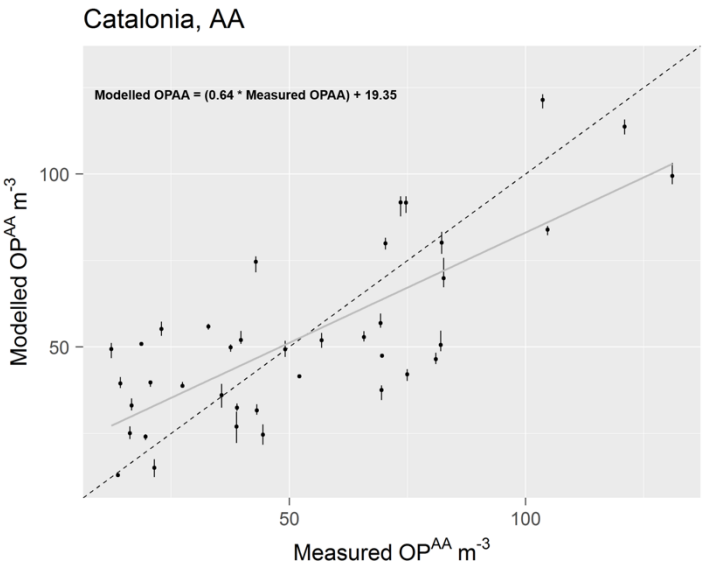
Figure S3. Scatterplots of hold-out validation (HOV) by study area (dashed lines are 1:1; solid lines are linear regression slopes; bars represent the range of predictions for each monitoring site; dots represent the median prediction for each monitoring site).

Basel, Switzerland



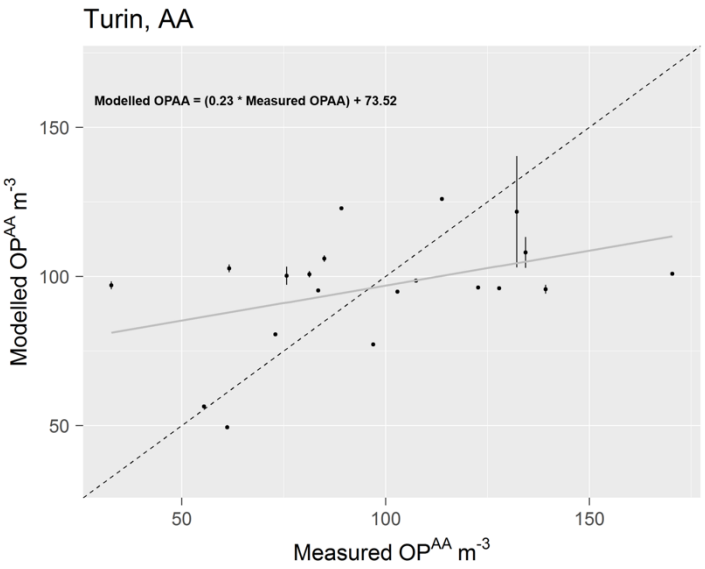
GSH – no model

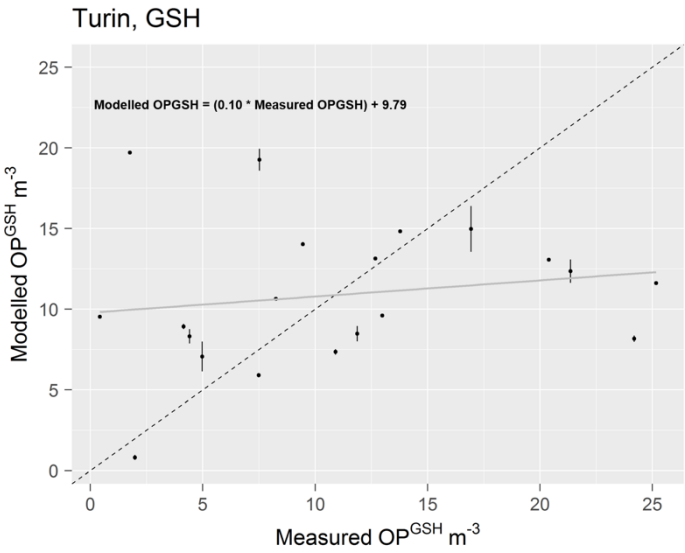
Catalonia, Spain



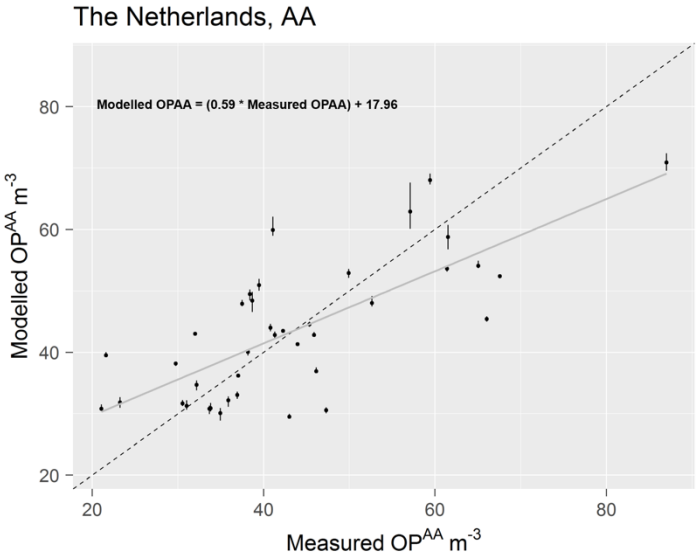
GSH – model not valid following RSS

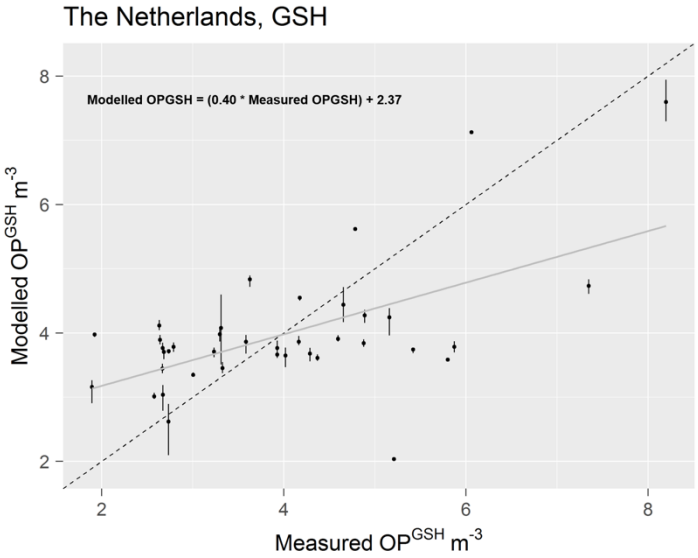
Turin, Italy


Turin, AA
Modelled OPAA = (0.23 * Measured OPAA) + 73.52
Modelled OP^AA m^-3
Measured OP^AA m^-3
50
100
150

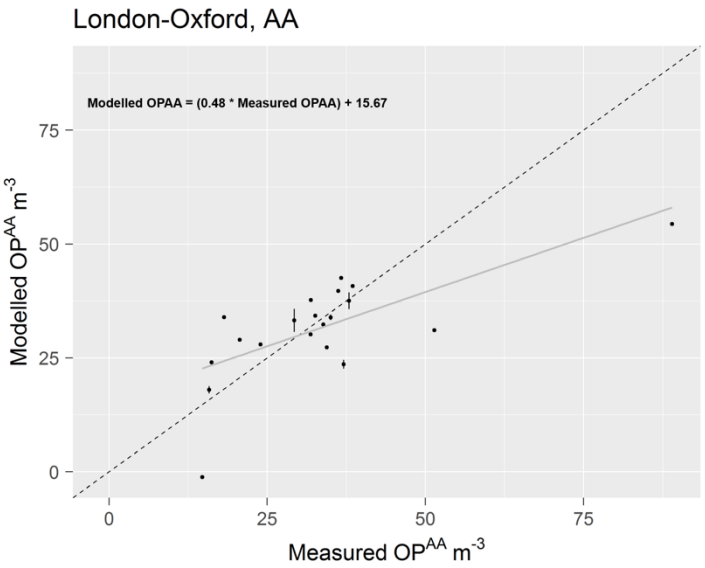

Turin, GSH
Modelled OPGSH = (0.10 * Measured OPGSH) + 9.79
Modelled OP^GSH m^-3
Measured OP^GSH m^-3
0
5
10
15
20
25

The Netherlands



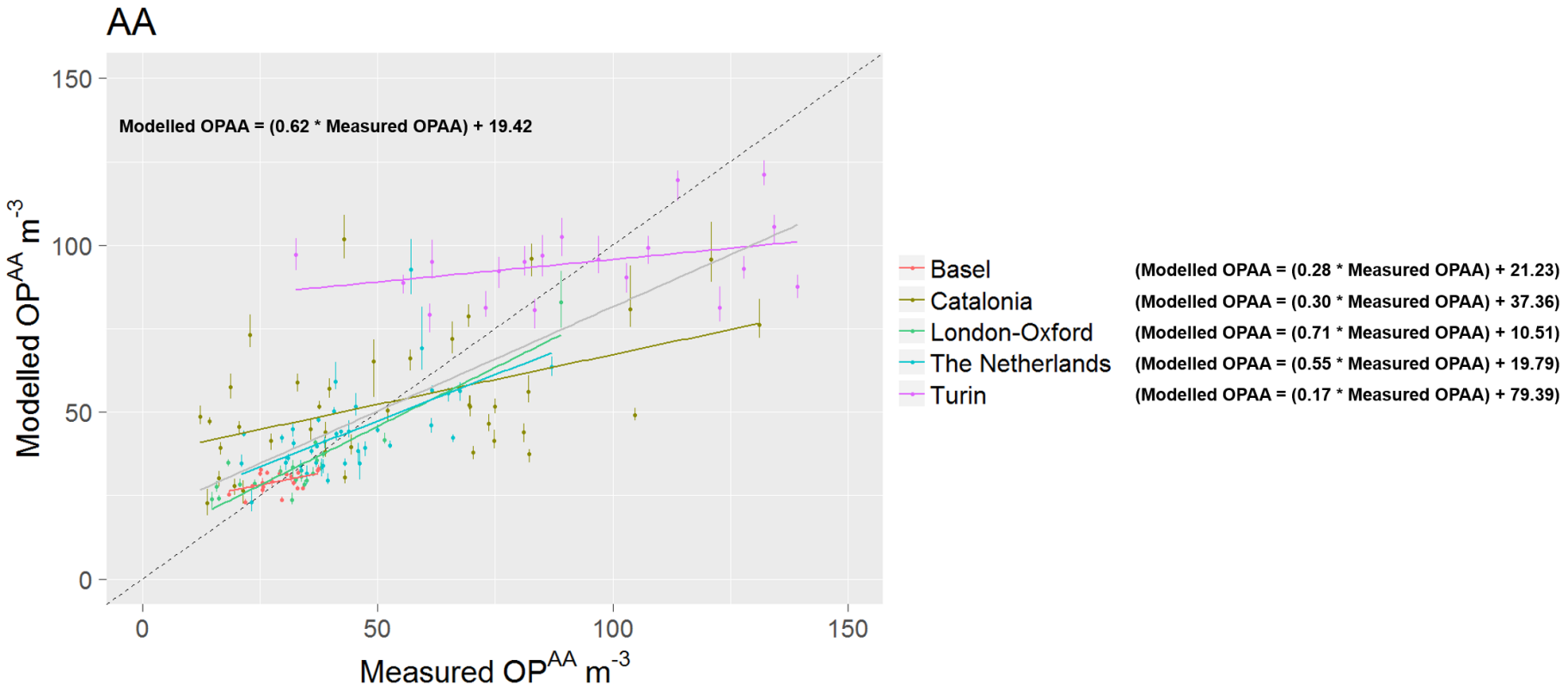


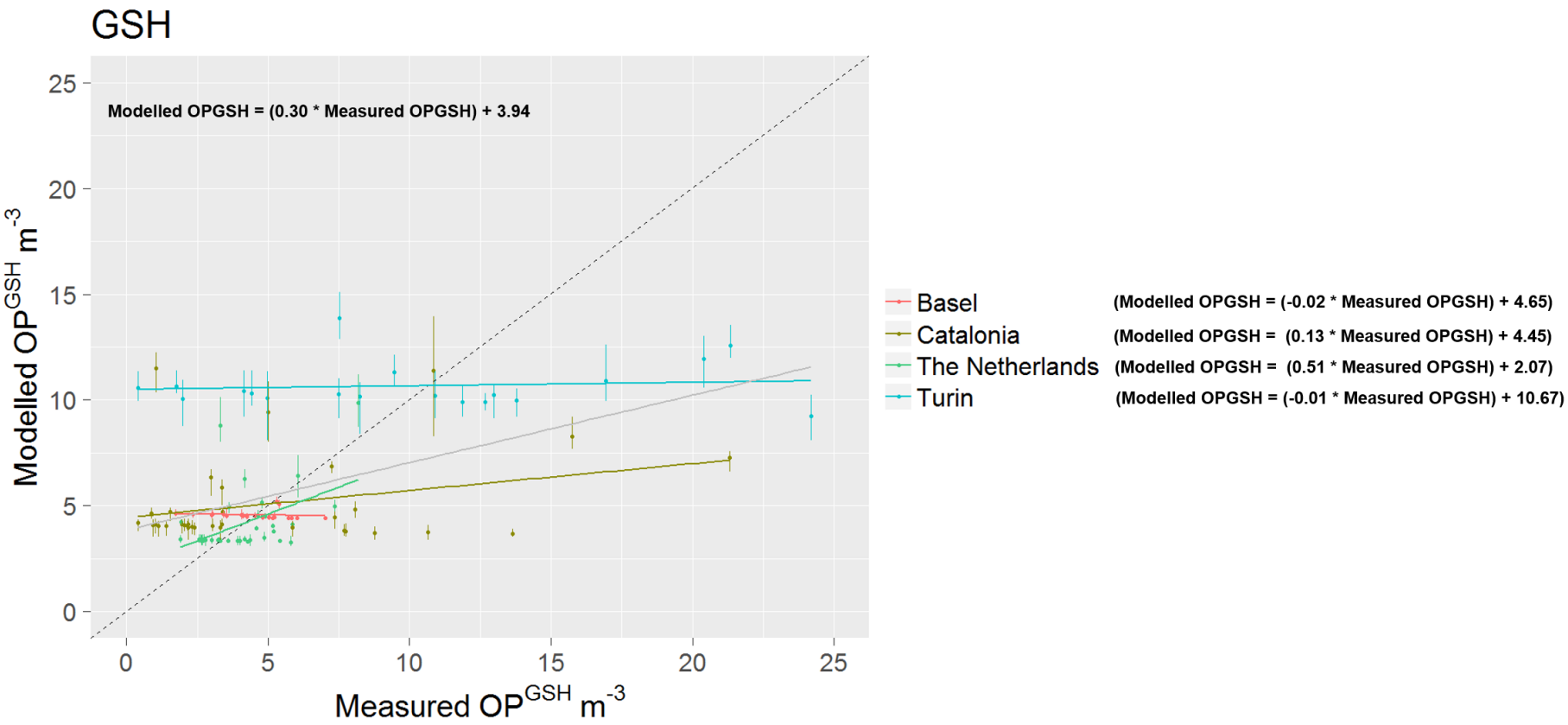
London-Oxford, United Kingdom



GSH – no measurements to derive a model

Combined models (all areas)




GSH
Modelled OPGSH = (0.30 * Measured OPGSH) + 3.94
Modelled OP^GSH m^-3
Measured OP^GSH m^-3
0
5
10
15
20
25
Basel (Modelled OPGSH = (-0.02 * Measured OPGSH) + 4.65)
Catalonia (Modelled OPGSH = (0.13 * Measured OPGSH) + 4.45)
The Netherlands (Modelled OPGSH = (0.51 * Measured OPGSH) + 2.07)
Turin (Modelled OPGSH = (-0.01 * Measured OPGSH) + 10.67)